Identification analysis of DSGE models*

Nikolay Iskrev

Abstract

A fully-specified DSGE model provides a complete characterization of a data generating process. Therefore, properties such as parameter identification are determined by the features of the underlying economic model. This paper shows how to analyze the identification properties of DSGE models by addressing the following questions: Are parameters identified and how well? What are the causes of identification problems? Which are the main sources of information about individual parameters? How does identification change across the parameter space, with the set of observables and the sample size? The analysis exploits the properties of the log-likelihood function, does not require actual data or estimation, and is simple to perform for large scale DSGE model. The methodology is illustrated using the medium-scale DSGE model estimated in Smets and Wouters (2007).

Keywords: DSGE models, Identification, Information matrix, Cramér-Rao lower bound

JEL classification: C32, C51, C52, E32

*This is a substantially revised version of Chapter 3 of my dissertation at the University of Michigan. A version of the paper previously circulated under the title "Evaluating the strength of identification in DSGE models. An a priori approach". I am grateful to my dissertation committee of Robert Barsky, Fred Feinberg, Christopher House, and Serena Ng for their guidance. I also thank Alejandro Justiniano, Frank Schorfheide, and Marco Ratto for helpful comments and discussions.

Contact Information: Banco de Portugal, Av. Almirante Reis 71, 1150-012, Lisboa, Portugal
e-mail: Nikolay.Iskrev@bportugal.pt; phone:+351 213130006

1 Introduction

There is a considerable consensus among academic economists and economic policy makers that modern macroeconomic models are rich enough to be useful as tools for policy analysis. It is also well understood that when structural models are used for quantitative analysis, it is critical to use parameter values that are empirically relevant. The best way to obtain such values is to estimate and evaluate the models in a formal and internally consistent manner. This is what the empirical dynamic stochastic general equilibrium (DSGE) literature attempts to do.

The estimation of DSGE models exploits the restrictions they impose on the joint probability distribution of observed macroeconomic variables. A fundamental question that arises is whether these restrictions are sufficient to allow reliable estimation of the model parameters. This is known in econometrics as the identification problem. To answer it, econometricians study the relationship between the true probability distribution of the data and the parameters of the underlying economic model (Koopmans (1949)). Such identification analysis should precede the statistical estimation of economic models (Manski (1995)).

Although the importance of parameter identification has been recognized, the issue is rarely discussed when DSGE are estimated. Examples of models with unidentifiable parameters can be found in Kim (2003), Beyer and Farmer (2004) and Cochrane (2007). That DSGE models may be poorly identified has been pointed out by Sargent (1976) and Pesaran (1989). More recently, Canova and Sala (2009) summarize their study of identification issues in DSGE models with the conclusion that: "it appears that a large class of popular DSGE structures can only be weakly identified".

Most of the existing research on identification in DSGE models follows the econometric literature in which weak identification is treated as a sampling problem, i.e. as something within the realm of statistical inference (see e.g. Stock and Yogo (2005) and the survey in Andrews and Stock (2005)). For this reason, the effort has been devoted to either devising tests for detecting weak identification (Inoue and Rossi (2011)), or to developing methods for inference that are robust to identification problems (e.g. Guerron-Quintana et al. (2013), Qu (2014) and Andrews and Mikusheva (2014)).

This paper pursues a different approach, whereby parameter identification is treated as a property of the underlying economic model. In contrast to other types of models, where the mapping from economic model to data is only partially known, DSGE models provide a complete characterization of a data generating process. Thus, identification problems that may appear in a particular data set must have their origins in the underlying structural model. Identification problems would occur when the restrictions the model imposes on

the joint distribution of the observed variables are not sufficiently informative about some parameters. In that sense identification is a property of the model. However, since the information content of the restrictions is, in general, a function of the parameter values, the data also plays a role in providing parameter values for which the model mimics the behavior of the observed variables.

The purpose of this paper is to show how to conduct identification analysis of DSGE models. The particular questions of interest include: (1) which model parameters are identified and which are not; (2) how well identified the identifiable parameters are; (3) what are the causes for identification problems; (4) what are the main sources of identification; and (5) how the answers to (1)-(4) change across regions in the parameter space and across different sets of observables or sample sizes.

A central tool in the proposed approach is the expected Fisher information matrix, the use of which for identification analysis was first suggested by Rothenberg (1971). The information matrix measures the curvature of the expected log-likelihood surface and, as Rothenberg points out, it "is a measure of the amount of information about the unknown parameters available in the sample". Identification problems arise when the log-likelihood surface is flat or nearly flat with respect to some parameters. This can be detected and quantified with the help of the information matrix. Furthermore, a decomposition of the matrix can be used to determine the roots of the identification problems. Parameters would be unidentifiable or weakly identified if the economic features they represent are nearly or completely irrelevant with respect to the variables used to estimate the model. This may occur either because those features are unimportant on their own, or because they are nearly redundant given other features represented in the model. These issues are particularly relevant for DSGE models, which are sometimes criticized for being too rich in features, and possibly overparameterized (Chari et al. (2009)).

Papers related to this one are Iskrev (2010), Komunjer and Ng (2011), Qu and Tkachenko (2012b) and Koop et al. (2013), which consider the parameter identifiability question, and Canova and Sala (2009), which focuses on the weak identification problem. Iskrev (2010) presents an identifiability condition that is easier to use and more general than the one developed here. The condition is based on the Jacobian matrix of the mapping from theoretical first and second order moments of the observable variables to the deep parameters of the model. The condition is necessary and sufficient for identification with likelihood-based methods under normality, or with limited information methods that utilize only first and second order moments of the data. However, that paper does not address the weak identification issue, which is one of the main themes of this paper. Komunjer and Ng (2011) derive a similar rank condition for identification using the spectral density matrix, while Qu and Tkachenko (2012b)

add a condition for identification from a subset of frequencies. Koop et al. (2013) consider parameter identification in DSGE models from a Bayesian perspective. Canova and Sala (2009) were the first to draw attention to the problem of weak identification in DSGE models, as well as discuss different strategies for detecting it. Those include: one and two dimensional plots of the estimation objective function, estimation with simulated data, and checking numerically the conditioning of matrices characterizing the mapping from parameters to the objective function. Canova and Sala (2009) differs from the present paper in several ways. First, they approach parameter identification from the perspective of a particular limited information estimation method, namely, equally weighted impulse response matching. In addition to the model and data deficiencies discussed above, weak identification in that setting may be caused by the failure to use some model-implied restrictions on the distribution of the data, and by the inefficient weighing of the utilized restrictions. Consequently, it may be very difficult to disentangle the causes and quantify their separate contribution to the identification problems. Second, it is very common in DSGE models to have identification problems that stem from a near observational equivalence involving a large number of parameters. This means that the objective function is flat with respect to all of the parameters as a group. The plots used in Canova and Sala (2009) are limited to only two parameters at a time, and it is far from straightforward to select the appropriate pairs from a large number of free parameters. Third, Canova and Sala (2009) do not discuss the role of the set of observables for identification. The effect of using different observables for the estimation of a DSGE model is investigated in Guerron-Quintana (2010) who finds that the parameter estimates and the economic and forecasting implications of the model vary substantially with the choice of included variables. The last and perhaps most important difference is in the approach itself. A key advantage of the method described here is that the expected information matrix can be evaluated analytically for linear Gaussian models. As a result, identification analysis can be performed easily even for large-scale models under different assumptions about the parameter values, the sample size, the set of observed variables, the choice of parameters to be calibrated instead of estimated, etc. While it is, in principle, possible to address these questions by conducting Monte Carlo simulations, this is hardly a viable strategy for most DSGE models. Estimating a multidimensional and highly non-linear model even once is a numerically challenging and time consuming exercise. Attempting this many times under different assumptions about the parameters, the sample size, and the observables would be impractical.

Another important aspect of the identification analysis concerns the sources of identification of the parameters. One way to think about this question is in terms of the moments of the data, which, according to the model, should be most informative about individual parameters. In Iskrev (2014), this is analyzed using the weights assigned to moments in the

first order conditions of the general method of moments estimator with optimal weighting matrix. In the present paper, it is shown how to extend that analysis to likelihood based estimation, where the weights assigned to moments are obtained from the score vector. In addition, the question is approached from a frequency domain perspective by asking what parts of the spectrum are most informative about the parameters of the model.

The remainder of the paper is organized as follows. Section 2 introduces the class of linearized DSGE models, and outlines the derivation of the log-likelihood function and the Fisher information matrix for Gaussian models. Section 3 explains the role of the Fisher information matrix in the analysis of identification, and describes how to evaluate and analyze the strength of identification and how to determine the sources of identification from the score. The methodology is illustrated, in Section 4, with the help of the medium-scale DSGE model estimated in Smets and Wouters (2007). Concluding comments are given in Section 5.

2 Preliminaries

This section provides a brief discussion of the class of linearized DSGE models as well as the derivation of the log-likelihood function and the Fisher information matrix for Gaussian models.

2.1 Setup

I consider DSGE models expressed in terms of stationary variables and linearized around the steady state values of these variables. Such models can be expressed as follows:

$$\boldsymbol{\Gamma}_0(\boldsymbol{\theta})\boldsymbol{z}_t = \boldsymbol{\Gamma}_1(\boldsymbol{\theta})\,\mathrm{E}_t\,\boldsymbol{z}_{t+1} + \boldsymbol{\Gamma}_2(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_3(\boldsymbol{\theta})\boldsymbol{\epsilon}_t \tag{2.1}$$

where $\boldsymbol{z}_t$ is an $m-$dimensional vector of deviations from steady states, and $\boldsymbol{\epsilon}_t$ is an n-dimensional random vector of structural shocks with $\boldsymbol{\epsilon}_t \sim i.i.d.\ \mathbb{N}(\mathbf{0}, \boldsymbol{I}_n)$. The elements of the matrices $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\Gamma}_3$ are functions of a $k-$dimensional vector of deep parameters $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a point in $\boldsymbol{\Theta} \subset \mathbb{R}^k$. The parameter space $\boldsymbol{\Theta}$ is defined as the set of all theoretically admissible values of $\boldsymbol{\theta}$.

The solution of equation (2.1) can be expressed as a linear state space model:

$$\boldsymbol{x}_t = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{z}_t \tag{2.2}$$

$$\boldsymbol{z}_t = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t \tag{2.3}$$

where $\boldsymbol{x}_t$ is a l-dimensional vector of observed variables, $\boldsymbol{s}(\boldsymbol{\theta})$ is a l-dimensional vector, $\boldsymbol{C}$ is

a $l \times m$ matrix, $\boldsymbol{A}$ is a $m \times m$ matrix, and the $\boldsymbol{B}$ is a $m \times n$ matrix.

Remark 1. It is straightforward to introduce measurement errors into the system (2.2)-(2.3) by expanding the vectors $\boldsymbol{z}_t$ and $\boldsymbol{\epsilon}_t$ and making the necessary changes in the state space matrices.

2.2 Log-likelihood function and the information matrix

The log-likelihood function of the data $\boldsymbol{X}_T = [\boldsymbol{x}_1', \ldots, \boldsymbol{x}_{\mathrm{T}}']'$ can be constructed using the prediction error method whereby a sequence of one-step ahead prediction errors, $\boldsymbol{e}_{t|t-1} = \boldsymbol{x}_t - \boldsymbol{s} - \boldsymbol{C}\hat{\boldsymbol{z}}_{t|t-1}$, is constructed by applying the Kalman filter to obtain one-step ahead forecasts of the state vector $\hat{\boldsymbol{z}}_{t|t-1}$. The Gaussianity of the structural shocks implies that the conditional distribution of $\boldsymbol{e}_{t|t-1}$ is also Gaussian with mean zero and a covariance matrix given by $\boldsymbol{S}_{t|t-1} = \boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}'$, where $\boldsymbol{P}_{t|t-1} = \mathrm{E}\left(\boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1}\right)\left(\boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1}\right)'$ is the conditional covariance matrix of the one-step ahead forecast, and is also obtained from the Kalman filter recursion. This implies that the log-likelihood function of the sample is given by:

$$\ell_T(\boldsymbol{\theta}) = \text{const.} - \frac{1}{2}\sum_{t=1}^{\mathrm{T}} \log(|\boldsymbol{S}_{t|t-1}|) - \frac{1}{2}\sum_{t=1}^{\mathrm{T}} \boldsymbol{e}_{t|t-1}^{'}\boldsymbol{S}_{t|t-1}^{-1}\boldsymbol{e}_{t|t-1} \tag{2.4}$$

Under some regularity conditions, the maximum likelihood estimator $\tilde{\boldsymbol{\theta}}_T$ is consistent, asymptotically efficient and asymptotically normally distributed with:

$$\sqrt{\mathrm{T}}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbb{N}\left(\boldsymbol{0}, \boldsymbol{\mathcal{I}}_0^{-1}\right) \tag{2.5}$$

Here $\boldsymbol{\mathcal{I}}_0$ is the asymptotic Fisher information matrix evaluated at the true value of $\boldsymbol{\theta}$. That is

$$\boldsymbol{\mathcal{I}}_0 := \lim_{T\to\infty}\left(\frac{1}{T}\boldsymbol{\mathcal{I}}_T\right) \tag{2.6}$$

where $\boldsymbol{\mathcal{I}}_T$ is the finite sample Fisher information matrix, defined as:

$$\boldsymbol{\mathcal{I}}_T := \mathrm{E}\left[\left\{\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right\}' \left\{\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right\}\right] \tag{2.7}$$

The computation of the asymptotic and the finite sample information matrices for Gaussian linear state space models is discussed, among others, in Zadrozny (1989), Segal and Weinstein (1989), Zadrozny and Mittnik (1994), and Klein and Neudecker (2000).

3 Identification Analysis

This section gives an overview of the concept of identification in econometrics, and shows how to measure the strength of identification, analyze the causes for identification problems and determine the main sources of identification of the parameters in the model from Section 2.

3.1 General principles

Let a model be parameterized in terms of a vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$ and suppose that inference about $\boldsymbol{\theta}$ is made on the basis of T observations of a random vector $\boldsymbol{x}$ with a known joint probability density function $p(\boldsymbol{X}_T; \boldsymbol{\theta})$, where $\boldsymbol{X}_T = [\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T']'$. When considered as a function of $\boldsymbol{\theta}$, $p(\boldsymbol{X}_T; \boldsymbol{\theta})$ contains all available sample information about the value of $\boldsymbol{\theta}$ associated with the observed data. Thus, a basic prerequisite for making inference about $\boldsymbol{\theta}$ is that distinct values of $\boldsymbol{\theta}$ imply distinct values of the density function. Formally, we say that a point $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is identified if:

$$p(\boldsymbol{X}_T; \boldsymbol{\theta}) = p(\boldsymbol{X}_T; \boldsymbol{\theta}_0) \text{ with probability } 1 \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0 \tag{3.1}$$

This definition is made operational by using the following property of the log-likelihood function $\ell_T(\boldsymbol{\theta}) := \log p(\boldsymbol{X}_T; \boldsymbol{\theta})$:

$$\mathrm{E}_0\, \ell_T(\boldsymbol{\theta}_0) \geq \mathrm{E}_0\, \ell_T(\boldsymbol{\theta}), \text{ for any } \boldsymbol{\theta} \tag{3.2}$$

This follows from the Jensen's inequality (see Rao (1973)) and the fact that the logarithm is a concave function. It further implies that the function $H(\boldsymbol{\theta}_0, \boldsymbol{\theta}) := \mathrm{E}_0\left(\ell_T(\boldsymbol{\theta}) - \ell_T(\boldsymbol{\theta}_0)\right)$ achieves a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and $\boldsymbol{\theta}_0$ is identified if and only if that maximum is unique. While conditions for global uniqueness are difficult to find in general, local uniqueness of the maximum at $\boldsymbol{\theta}_0$ may be established by verifying the usual first and second order conditions, namely: (a) $\frac{\partial H(\boldsymbol{\theta}_0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$, (b) $\frac{\partial^2 H(\boldsymbol{\theta}_0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is negative definite. If the maximum at $\boldsymbol{\theta}_0$ is locally unique, we say that $\boldsymbol{\theta}_0$ is locally identified. This means that there exists an open neighborhood of $\boldsymbol{\theta}_0$ where (3.1) holds for all $\boldsymbol{\theta}$. Global identification, on the other hand, extends the uniqueness of $p(\boldsymbol{X}_T; \boldsymbol{\theta}_0)$ to the whole parameter space. One can show that (see Bowden (1973)) the condition in (a) is always true, and the Hessian matrix in (b) is equal to the negative of the Fisher information matrix. Thus, we have the following result of Rothenberg (1971):

Theorem 1. *Let $\boldsymbol{\theta}_0$ be a regular point of the information matrix $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta})$. Then $\boldsymbol{\theta}_0$ is locally identifiable if and only if $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}_0)$ is non-singular.*

A point is called regular if it belongs to an open neighborhood where the rank of the

matrix does not change. Without this assumption, the condition is only sufficient for local identification. Although it is possible to construct examples where regularity does not hold (see Shapiro and Browne (1983)), typically the set of irregular points is of measure zero (see Bekker and Pollock (1986)). Thus, for most models the non-singularity of the information matrix is both necessary and sufficient for local identification.

Remark 2. Note that the Rothenberg condition for local identification involves the finite sample information matrix $\boldsymbol{\mathcal{I}}_T$. Non-singularity of the asymptotic information matrix $\boldsymbol{\mathcal{I}}_0$ is a condition for asymptotic local identification of the parameters, which means that $\operatorname{plim} \frac{1}{T}\ell(\boldsymbol{\theta}) \neq \operatorname{plim} \frac{1}{T}\ell(\boldsymbol{\theta}_0)$ for all $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}_0$ such that $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Asymptotic identification is necessary but not sufficient for identification in the sense discussed here.

Remark 3. When the probability density function is a member of the exponential family, the non-singularity of the information matrix can be established by checking the rank of a Jacobian matrix, which is often easier to compute (see Wansbeek and Meijer (2000)). In the case of a multivariate normal distribution, the Jacobian matrix is constructed using the first order derivatives of the first and second order moments of the variables. This condition is used in Iskrev (2010) to check for identification in DSGE models.

Verifying that the model is identified, at least locally, is important since identifiability is a prerequisite for the consistent estimation of the parameters. Singularity of the information matrix means that the expected log-likelihood function is flat at $\boldsymbol{\theta}_0$ and one has no hope of finding the true values of some of the parameters even with an infinite number of observations. Intuitively, this may occur for one of two reasons. Either some parameters do not affect the expected log-likelihood at all, or different parameters have the same effect on the expected log-likelihood. This reasoning may be formalized by using the fact that the information matrix is equal to the covariance matrix of the scores, and therefore can be expressed as:

$$\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}_0) = \boldsymbol{\Delta}^{\frac{1}{2}} \boldsymbol{\mathcal{R}}_T(\boldsymbol{\theta}_0) \boldsymbol{\Delta}^{\frac{1}{2}} \tag{3.3}$$

where $\boldsymbol{\Delta} = \operatorname{diag}(\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}_0))$ is a diagonal matrix containing the variances of the elements of the score vector, and $\boldsymbol{\mathcal{R}}_T(\boldsymbol{\theta}_0)$ is the correlation matrix of the score vector.

Hence, a parameter θ_i is locally unidentifiable if:

(a) Small changes in θ_i have no effect on the expected log-likelihood, i.e.

$$\Delta_i := \mathrm{E}\left(\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i}\right)^2 = -\,\mathrm{E}\left(\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \theta_i^2}\right) = 0 \tag{3.4}$$

(b) The effect on the expected log-likelihood of small changes in θ_i can be offset by changing

other parameters, i.e.

$$\boldsymbol{\varrho}_i := \sqrt{1 - 1/\boldsymbol{\mathcal{R}}_T^{ii}} = 1, \tag{3.5}$$

where $\boldsymbol{\mathcal{R}}_T^{ii}$ is the i-th diagonal element of the inverse of $\boldsymbol{\mathcal{R}}_T$. The intuition about the meaning of $\boldsymbol{\varrho}_i$ comes from a well-known property of the correlation matrix (see e.g. Tucker et al. (1972)), which imply that $\boldsymbol{\varrho}_i$ is the coefficient of multiple correlation between the partial derivative of the log-likelihood with respect to θ_i and the partial derivatives of the log-likelihood with respect to the other elements of $\boldsymbol{\theta}$. Both (a) and (b) result in a flat expected log-likelihood function and lack of identification for one or more parameters. Weak identification, on the other hand, arises when the expected log-likelihood is not completely flat but exhibits very low curvature with respect to some parameters. The issue of detecting and measuring weak identification problems is discussed next.

3.2 Identification strength

The rank condition ensures that the expected log-likelihood function is not flat and achieves a locally unique maximum at the true value of $\boldsymbol{\theta}$. In general, this suffices for a consistent estimation of $\boldsymbol{\theta}$. However, the precision with which $\boldsymbol{\theta}$ may be estimated in finite samples depends on the degree of curvature of the expected log-likelihood surface in the neighborhood of $\boldsymbol{\theta}_0$, of which the rank condition provides no information. Nearly flat expected log-likelihood means that small changes in the value of $\ell_T(\boldsymbol{\theta})$, due to random variations in the sample, could result in very large changes in the value of $\boldsymbol{\theta}$ that maximizes the observed likelihood function. When this occurs, parameter identification is said to be weak in the sense that the estimates are prone to be very inaccurate even when the number of observations is large. In other words, a parameter is weakly identified if the degree of precision with which it can be estimated with a sample of a given size is unacceptably low. In that sense, what "weak" means depends on what is considered unacceptable, and is therefore a relative not an absolute concept.

As explained in Rothenberg (1971), the curvature of the expected log-likelihood function is described by the Fisher information matrix. The relationship between the curvature and the precision of the ML estimator $\hat{\boldsymbol{\theta}}_T$ can be seen from the asymptotic distribution of the latter. In particular, (2.5) implies that $\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}_0)/T$ is an approximation of the sampling covariance matrix of $\hat{\boldsymbol{\theta}}_{\mathrm{T}}$, and $\boldsymbol{\mathcal{I}}^{ii}(\boldsymbol{\theta}_0)/T$ approximates the sampling variance of $\hat{\theta}_i$, where $\boldsymbol{\mathcal{I}}^{ii}$ is the i-th diagonal element of the inverse of the information matrix. The asymptotic normality of $\hat{\boldsymbol{\theta}}_T$ may also be used to construct asymptotic joint confidence sets for $\boldsymbol{\theta}$ as a whole and asymptotic confidence intervals for each θ_i. However, these asymptotic results might be unreliable in finite samples. Specifically, to accurately characterize the uncertainty about an estimate, one has to take into account the full shape of the log-likelihood function of the sample. In contrast,

the asymptotic confidence intervals are constructed on the basis of a quadratic approximation of the expected log-likelihood whose shape is represented by the curvature. The two types of intervals may be quite different when the log-likelihood function is far from quadratic. At the same time, for reasonably smooth functions, the curvature of the log-likelihood function would be an informative indicator of whether a parameter is well identified or not.

The asymptotic efficiency of MLE means that the estimator has the smallest asymptotic covariance matrix among all consistent estimators. This follows from the Cramér-Rao theorem which states that the asymptotic covariance of any consistent estimator of $\boldsymbol{\theta}$ is bounded from below by the inverse of the asymptotic information matrix, $\boldsymbol{\mathcal{I}}_0$. On the other hand, the inverse of the finite sample information matrix $\boldsymbol{\mathcal{I}}_T$ is a lower bound on the covariance matrix of any unbiased estimator. This implies that $\mathrm{b}_i := \boldsymbol{\mathcal{I}}_T^{ii}$ is a lower bound on the variance of any unbiased estimator of θ_i and can be used to measure the strength of identification of individual parameters in terms of bounds on one-standard-deviation intervals for the parameters. There exists a direct relationship between the size of the bounds and the possible causes of identification problems. Using the decomposition of $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta})$ shown in (3.3) and the properties of the correlation matrix, it is easy to show that the following relation holds:

$$\mathrm{b}_i = \frac{1}{\Delta_i(1-\boldsymbol{\varrho}_i^2)} \tag{3.6}$$

Thus, b_i may be large either because $\Delta_i \approx 0$ or because $\boldsymbol{\varrho}_i \approx 1$. In the first case, the parameter is nearly irrelevant as it has only a weak effect on the likelihood. In the second case, it is nearly redundant because its effect on the likelihood can be approximated very well by other parameters. Consequently, the value of that parameter will be difficult to pin down on the basis of information contained in the likelihood function.

It is worth highlighting that strong sensitivity of the likelihood with respect to a parameters is not a guarantee that the parameter is well identified. Note that $\Delta_i = \mathrm{E}\left(\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \theta_i}\right)^2$ is the inverse of the Cramér-Rao lower bound for θ_i, given that the other parameters, i.e. the elements of $\boldsymbol{\theta}_{-i}$, are known. Even if Δ_i is large, the identification of θ_i might be very weak if $\boldsymbol{\varrho}_i$ is close to 1. This observation clarifies the difference between the information in the likelihood about a parameter θ_i when the other parameters are known, given by Δ_i, and the information about θ_i when the other parameters are unknown, given by b_i. In general, the second type of information is smaller and the difference increases with the value of $\boldsymbol{\varrho}_i$.[1]

[1]The difference between the two types of information about θ_i can also be seen in terms of the difference between the expected curvature of the log-likelihood with respect to θ_i, and the expected curvature of the profile log-likelihood function of θ_i. The first is equal to i-th diagonal element of $\boldsymbol{\mathcal{I}}_T$, while the second is given by the inverse of i-th diagonal element of the inverse of $\boldsymbol{\mathcal{I}}_T$.

Example 1. Consider the following ARMA(1,1) model:

$$x_t = \phi_1 x_{t-1} + \epsilon_t - \phi_2 \epsilon_{t-1}, \qquad |\phi_1| < 1, |\phi_2| < 1, \qquad \epsilon_t \sim \mathbb{N}(0, \sigma^2) \tag{3.7}$$

For simplicity, assume that σ^2 is known. The information matrix for $\boldsymbol{\theta} := [\phi_1, \phi_2]'$ is

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{1-\phi_1^2} & \frac{-1}{1-\phi_1\phi_2} \\ \frac{-1}{1-\phi_1\phi_2} & \frac{1}{1-\phi_2^2} \end{bmatrix} \tag{3.8}$$

and the diagonal elements of the inverse of $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ are:

$$\mathrm{b}_i = \frac{(1-\phi_1\phi_2)^2(1-\phi_i^2)}{(\phi_1-\phi_2)^2} \qquad \text{for } i = 1, 2 \tag{3.9}$$

From (3.9), it is clear that ϕ_1 and ϕ_2 are not identified if $\phi_1 = \phi_2$ and that they are weakly identified when $\phi_1 \approx \phi_2$. Furthermore, we can express (3.9) in the form of (3.6) using $\Delta_i = 1/(1-\phi_i^2)$ and $\boldsymbol{\varrho}_i^2 = (1-\phi_1^2)(1-\phi_2^2)/(1-\phi_1\phi_2)^2$, which shows that the reason for weak identification is that $\boldsymbol{\varrho}_i \approx 1$ when $\phi_1 \approx \phi_2$. This implies that the effects of ϕ_1 and ϕ_2 on the likelihood are very similar and the two parameters are difficult to identify separately.

If the model is re-parametrized in terms of $\psi := \phi_1 - \phi_2$ and ϕ_2, we have

$$\mathrm{b}_i = \frac{(1-\phi_2^2)(1-\phi_2^2-\phi_2\psi)}{\psi^2} \tag{3.10}$$

Therefore, ϕ_2 is unidentified when $\psi = 0$ and weakly identified when $\psi \approx 0$. Again, b_i can be decomposed as in (3.6) using $\Delta_i = \psi^2(1+\psi\phi_2+\phi_2^2)/(1-\phi_2^2)(1-(\psi+\phi_2)^2)(1-\psi\phi_2-\phi_2^2)$ and $\boldsymbol{\varrho}_2^2 = (\psi+\phi_2)^2(1-\phi_2^2)/\left(1-(\phi_2^2+\gamma\phi_2)^2\right)$. Now the cause for weak identification is that $\Delta_i \approx 0$ when $\psi \approx 0$, which means that the effect of ϕ_2 on the likelihood is very weak and vanishes in the limit.

The problem with identification in the ARMA(1,1) model with near cancelling roots is well known and the results presented above are not new. In this (and other) simple models parameter identification problems can be analyzed directly, without the use of the information matrix. In the much larger and complicated DSGE models, however, this is not feasible since the relationship between the structural parameters and the reduced form representation is typically not available in explicit analytical form. Therefore, the information matrix and the decomposition in (3.6), which can easily be evaluated for such models, could be very useful tools for detecting problems with identification and understanding the causes for these problems.

3.3 Sources of identification

To evaluate the strength of identification of the parameters, we take into account all model-implied restrictions on the joint probability distribution of the observables. A natural next step in the analysis is to ask which characteristics of the distribution are most important for the identification of individual parameters. In principle, one may be able to answer this question by reasoning alone, i.e. by tracing the link between the economic features represented by the parameters, and the properties of the data the model is designed to explain. In practice, however, the relationship between parameters and empirical implications of a model is often difficult to discern, especially for large models. It is therefore useful to have a formal method for doing that. In this section, I present two complementary approaches. The first one uses the first order conditions of the MLE to rank the moments of the observed variables in terms of their informativeness about the parameters of the model. The second approach uses the frequency domain approximation of the information matrix to compare the amount of information about individual parameters contained in different parts of the spectrum.

Most informative moments

The idea, in a nutshell, is this. Since the shocks are Gaussian, the distribution of the observables is completely characterized by the first and second order moments of the variables. Under correct model specification, the maximum likelihood estimator uses information contained in these moments efficiently so as to achieve the smallest asymptotic covariance matrix among all consistent estimators. To rank the moments in terms of their informativeness about the parameters, we ask how individual moments are weighted by the maximum likelihood estimator.

The maximum likelihood estimator is defined as the value of $\boldsymbol{\theta}$ that maximizes the log-likelihood function. Alternatively, it can be interpreted as the value of $\boldsymbol{\theta}$ which solves the system of first order conditions:

$$\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \mathbf{0}$$

To see how moments of $\boldsymbol{x}_t$ enter into this system, consider first the case when both the data and the model are demeaned. In the Appendix it is shown that, when $\boldsymbol{s}(\boldsymbol{\theta}) = \mathbf{0}$, the state space system (2.2)-(2.3) can be written as follows:

$$\boldsymbol{X}_T = \boldsymbol{L}\boldsymbol{E}_T \tag{3.11}$$

where $\boldsymbol{L}$ is a lower triangular matrix with unity diagonal elements and $\boldsymbol{E}_T := \left[\boldsymbol{e}'_{1|0}, \boldsymbol{e}'_{2|1}, \ldots, \boldsymbol{e}'_{T|T-1}\right]'$.

Since $\boldsymbol{E}_T$ is jointly Normal with covariance matrix $\boldsymbol{S} := \text{diag}([\boldsymbol{S}_{1|0}, \ldots, \boldsymbol{S}_{T|T-1}])$, we have

$$\boldsymbol{X}_T \sim \mathbb{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right), \text{ where } \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{L}\boldsymbol{S}\boldsymbol{L}' \tag{3.12}$$

Therefore, the log-likelihood function of $\boldsymbol{X}_T$ is given by:

$$\ell_T(\boldsymbol{\theta}) = const. - \frac{1}{2}\log\det\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{X}_T'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}_T \tag{3.13}$$

Differentiating (3.13) with respect to θ_i yields:

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_i} = \frac{1}{2}\text{vec}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)'\text{vec}\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) \tag{3.14}$$

where $\widehat{\boldsymbol{\Sigma}} := \boldsymbol{X}_T\boldsymbol{X}_T'$. Rearranging the first term of the product in (3.14) gives:

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_i} = \text{vec}\left(\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_i}\right)'\frac{1}{2}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\otimes\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)\text{vec}\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) \tag{3.15}$$

This expression for the score is the same as the first order conditions for a GMM estimator with moment conditions given by $\text{vec}\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) = \boldsymbol{0}$ and weighting matrix equal to $\frac{1}{2}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\otimes\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)$, which can be recognized as the information matrix of $\text{vec}\left(\widehat{\boldsymbol{\Sigma}}\right)$ (see e.g. Magnus and Abadir (2005)). Note that with l observed variables the matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is $lT \times lT$ and has $L^T := l^2(T-1) + l(l+1)/2$ unique elements representing various second order moments of $\boldsymbol{x}_t$ ($l(l+1)/2$ covariances and $l^2(T-1)$ autocovariances). If we let $m^j(\boldsymbol{\theta})$ be the j-th element of the L^T vector $\boldsymbol{m}(\boldsymbol{\theta})$ collecting these moments, and $\hat{m}_t^j$ be the sample realization of $m^j(\boldsymbol{\theta})$ at time t, we can write (3.15) as:

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_i} = \frac{1}{2}\sum_{j=1}^{L^T}\nu^{ij}\left(\hat{m}^j - m^j(\boldsymbol{\theta})\right) \tag{3.16}$$

where $\hat{m}^j := \dfrac{1}{\nu^{ij}}\sum_t \nu_t^{ij}\hat{m}_t^j$, $\nu^{ij} := \sum_t \nu_t^{ij}$, and ν_t^{ij} is equal to the sum of the elements of $\text{vec}\left(\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_i}\right)'\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\otimes\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)$ that multiply m_t^j.[2]

The expression in (3.16) shows that the MLE of the model (3.11) is indeed a moment matching estimator that minimizes the differences between theoretical second order moments

[2]Note that due to the symmetry of $\hat{\boldsymbol{\Sigma}}$, each off-diagonal element m_t^j appears in two positions in the vector $\text{vec}\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$.

and their sample counterparts. However, unlike the usual moment matching estimators, the empirical moments are not simple arithmetic averages, but are weighted averages where each data point receives a potentially different weight. The size, in absolute value, of the weights on the differences between empirical and theoretical moments determines how important it is to set those differences to zero. Since the scale of the moments in $\boldsymbol{m}(\boldsymbol{\theta})$ may be different, the weights on the moment conditions need to be multiplied by the values of the moments in order to comparable.[3] The relative importance of moment m^j for parameter θ_i is therefore measured by

$$\omega_{ij} = \frac{|\nu^{ij} m^j|}{\sum_{q=1}^{L^T} |\nu^{iq} m^q|} \tag{3.17}$$

Example 2. A simple model where the weights in the first order conditions of MLE can be studied analytically is the autoregressive of order one (AR(1)) process,

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \text{where } |\rho| < 1 \text{ and } \varepsilon_t \sim \mathbb{N}\left(0, \sigma^2\right) \tag{3.18}$$

The log-likelihood function of $\boldsymbol{X}_T := [x_1, x_2, \ldots, x_T]'$ can be written as a sequence of conditional distributions:

$$\begin{aligned} \ell(\rho, \sigma^2) &= log\left(f(x_1) f(x_2|x_1) f(x_3|, x_2) \ldots f(x_t|x_{t-1})\right) \\ &= -\frac{T}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\rho^2) - x_1^2\frac{(1-\phi^2)}{2\sigma^2} + \frac{1}{2\sigma^2}\sum_{t=2}^{T}(x_t - \rho x_{t-1})^2 \end{aligned} \tag{3.19}$$

The first order conditions with respect to ρ and σ^2 are:

$$\begin{aligned} \frac{\partial \ell}{\partial \rho} &= -\rho(T-2)\gamma(0)\frac{\left(\frac{1}{T-2}\sum_{t=2}^{T-1} x_t^2 - \gamma(0)\right)}{\gamma(0)} \\ &+\rho(T-1)\gamma(0)\frac{\left(\frac{1}{T-1}\sum_{t=2}^{T} x_t x_{t-1} - \gamma(1)\right)}{\gamma(1)} = 0 \end{aligned} \tag{3.20}$$

[3] Alternatively, the moment conditions could be normalized by the asymptotic standard deviations of the respective moments. This approach is preferable when some of the moments are zero or very close to zero. See Iskrev (2014) on how to compute analytically the required standard deviations

and

$$\frac{\partial \ell}{\partial \sigma^2} = \left(T + (T-2)\rho^2\right)\gamma(0)\frac{\left(\frac{x_1^2+x_T^2+(1+\rho^2)\sum_{t=2}^{T-1} x_t^2}{T+(T-2)\rho^2} - \gamma(0)\right)}{\gamma(0)} \tag{3.21}$$
$$-2\rho^2(T-1)\gamma(0)\frac{\left(\frac{1}{T-1}\sum_{t=2}^{T} x_t x_{t-1} - \gamma(1)\right)}{\gamma(0)} = 0$$

where $\gamma(h) := \mathrm{E}\, x_t x_{t-h}$ and I have use the fact that $\gamma(1) = \rho\gamma(0)$. Equations (3.20) and (3.21) reveal three important features of the maximum likelihood estimator: First, the estimator picks values of ρ and σ^2 which minimize the differences between the theoretical variance and first order autocovariance of x_t and their empirical counterparts. However, the empirical moments are not arithmetic averages, in which each realization is weighted equally, but instead are weighted averages of the sample realizations. Note that, in equation (3.20), x_1^2 and x_T^2 receive zero weights, while in (3.21), their weights are smaller than the weights on x_t for $2 \leq t \leq T-1$. Secondly, the maximum likelihood estimator does not use information in autocovariances beyond the first order even though they are available. In other words, the estimator assigns zero weights on terms such as $(x_t x_{t-h} - \gamma(h))$ for $2 \leq h \leq T-1$. Thirdly, in the first order condition for ρ, the relative weights on the two moment conditions are almost the same, except for very small values of T. In the first order condition for σ^2 the relative weight on first term is much larger for small values of ρ, and decreases to a half as ρ increases to 1. This implies that the two moment conditions are equally informative for ρ, while for σ^2 matching the empirical and theoretical variances is more important unless the process is very persistent.

In the general case, when the model and the data are not demeaned, $\boldsymbol{X}_T = \boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{L}\boldsymbol{E}_T$, with $\boldsymbol{\mu} := \iota_T \otimes \boldsymbol{s}(\boldsymbol{\theta})$ and ι_T being a T-dimensional vector of ones. The log-likelihood function of $\boldsymbol{X}_T$ is:

$$\ell(\boldsymbol{\theta}) = const. - \frac{1}{2}\log\det\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \frac{1}{2}\left(\boldsymbol{X}_T - \boldsymbol{\mu}(\boldsymbol{\theta})\right)'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\left(\boldsymbol{X}_T - \boldsymbol{\mu}(\boldsymbol{\theta})\right) \tag{3.22}$$

and the score is:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})'}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\left(\boldsymbol{X}_T - \boldsymbol{\mu}(\boldsymbol{\theta})\right) \tag{3.23}$$
$$+ \operatorname{vec}\left(\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right)'\frac{1}{2}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\otimes\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)\operatorname{vec}\left(\left(\boldsymbol{X}_T - \boldsymbol{\mu}(\boldsymbol{\theta})\right)\left(\boldsymbol{X}_T - \boldsymbol{\mu}(\boldsymbol{\theta})\right)' - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$$

The expression on the right-hand side of (3.23) can be put in the form of (3.16) with

two modifications: (1) the vector of moments $\boldsymbol{m}(\boldsymbol{\theta})$ now has $L^T + l$ elements, adding the mean of $\boldsymbol{x}_t$ to the second order moments; (2) the sample realizations of the second order moments are centred by subtracting the elements of $\boldsymbol{\mu}(\boldsymbol{\theta})$. Note that this, in addition to the potentially different weights on the sample realizations, is another difference between MLE and the standard GMM estimator, which weighs all realizations equally and centres them using the sample means.

Most informative frequencies

In addition to the time domain analysis, it may be of interest to know what are the sources of information in the frequency domain. Specifically, which part of the spectrum contains the most information about any given parameter. To answer this question, I use the fact that the Gaussian log-likelihood function can be approximated in the frequency domain as:[4]

$$\begin{aligned}\tilde{\ell}_T(\boldsymbol{\theta}) &= \text{const.} - \frac{1}{T}\sum_{j=0}^{T-1}\log\det(\boldsymbol{F}(\omega_j)) - \frac{1}{T}\sum_{j=0}^{T-1}\operatorname{tr}\left(\boldsymbol{F}(\omega_j)^{-1}\hat{\boldsymbol{F}}(\omega_j)\right) \\ &- \frac{T}{2}\operatorname{tr}\left[\boldsymbol{F}(0)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t - \boldsymbol{\mu}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t - \boldsymbol{\mu}\right)'\right] \qquad (3.24)\end{aligned}$$

where $\boldsymbol{F}(\omega) := \sum_{\tau=-\infty}^{\infty}\boldsymbol{\Sigma}(\tau)\exp(-i\omega\tau)$ is the spectral density matrix of $\boldsymbol{x}_t$ with $\boldsymbol{\Sigma}(\tau) := \mathrm{E}(\boldsymbol{x}_t - \boldsymbol{s}(\boldsymbol{\theta}))(\boldsymbol{x}_{t-\tau} - \boldsymbol{s}(\boldsymbol{\theta}))'$, $\hat{\boldsymbol{F}}(\omega_j) := \frac{1}{T}\boldsymbol{x}(\omega_j)\bar{\boldsymbol{x}}(\omega_j)'$ is the periodogram of $\boldsymbol{X}_T$ with the overbar denoting complex conjugation, and $\boldsymbol{x}(\omega_j) := \sum_{t=1}^{T}\boldsymbol{x}_t\exp(-i\omega_j t)$, $\omega_j = \frac{2\pi j}{T}$, $j = 0, \ldots, T-1$ is the Fourier transform of $\boldsymbol{X}_T$.

Papers estimating dynamic economic models by maximum likelihood in the frequency domain include Altug (1989), Diebold et al. (1998), Christiano and Vigfusson (2003). Qu and Tkachenko (2012a) and Sala (2014) estimate DSGE models using information from different frequency bands, and find that the parameter estimates and the economic implications of the model can differ substantially depending on which part of the spectrum is used. Qu and Tkachenko (2012b) show how to check whether the parameters of a DSGE model are identified from a subset of frequencies, while Qu (2014) considers identification robust inference in DSGE models from a frequency domain perspective.

The main argument for estimating a model in the frequency domain is the concern that the model may be misspecified with respect to certain frequencies. Fitting the model to only a subset of frequencies, e.g. business cycle frequencies, may alleviate this problem. The purpose of conducting identification analysis in the frequency domain is different, however. It is to understand what are the model implications regarding where in the spectrum is information

[4]See e.g. Hansen and Sargent (2014).

about different parameters concentrated. In other words, we are interested in the properties of the model as it is, and not whether it is a good model for explaining a particular data sample.[5]

Let $\tilde{\boldsymbol{\mathcal{I}}}_T(\boldsymbol{\theta})$ be the frequency domain approximation of the FIM, defined as in (2.7) with $\tilde{\ell}_T(\boldsymbol{\theta})$ instead of $\ell_T(\boldsymbol{\theta})$. Then the (u,v)-th element of $\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta})_T$ is given by[6]

$$\begin{aligned}\{\tilde{\boldsymbol{\mathcal{I}}}_T(\boldsymbol{\theta})\}_{u,v} &= \frac{T}{2\pi}\operatorname{tr}\left(\boldsymbol{F}(0)^{-1}\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_u}\right)\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_v}\right)'\right) \\ &+ \frac{1}{2}\sum_{j=0}^{T-1}\operatorname{tr}\left(\boldsymbol{F}(\omega_j)^{-1}\frac{\partial \boldsymbol{F}(\omega_j)}{\partial \theta_u}\boldsymbol{F}(\omega_j)^{-1}\frac{\partial \boldsymbol{F}(\omega_j)}{\partial \theta_v}\right) \qquad (3.25)\end{aligned}$$

The Appendix provides some details on how to evaluate that matrix for DSGE models. Note that both the log-likelihood (3.24) and the information matrix (3.25) are constructed by summing up terms which are independent across frequencies. Therefore, to evaluate the amount of information within a band of frequencies, one has to include the terms corresponding to these frequencies and omit the ones outside the band. As before, CRLBs for that set of frequencies are be obtained by inverting the resulting information matrix. The most informative part of the spectrum for a parameter is the one yielding the smallest value of the bound for that parameter.[7]

4 Application: Identification analysis of the Smets and Wouters (2007) model

In this section, I illustrate the identification analysis framework discussed above using a medium-scale DSGE model estimated in Smets and Wouters (2007) (SW07 henceforth). I start with an outline of the main components of the model, and then turn to the identification of the parameters.

[5]To help clarify this point, consider again the simple AR(1) process. It is easy to see that, the more persistent the process, the more volatility and therefore information there is in the low frequencies. This is a property of the AR(1) process irrespectively of whether it is a good description of any given sample of data.

[6]A classic reference on the frequency domain approximation of the FIM is Whittle (1953), who considers a large T approximation of the matrix for processes with zero mean. An extension of Whittle's formula to processes with non-zero mean is provided by Zeira and Nehorai (1990). See also Davies (1983) and the references therein.

[7]Alternatively, evaluating $\tilde{\boldsymbol{\mathcal{I}}}_T(\boldsymbol{\theta})$ for a large T yields an approximation of the asymptotic covariance matrix for MLE using the full spectrum. Summing over subsets of frequencies and inverting the corresponding matrices gives the asymptotic variances of the parameters. Comparing either the CRLBs or the asymptotic variances results in the same conclusions regarding the most informative parts of the spectrum.

4.1 The model

The model, based on the work of Smets and Wouters (2003) and Christiano et al. (2005), is an extension of the standard RBC model featuring a number of nominal frictions and real rigidities. These include monopolistic competition in goods and labor markets, sticky prices and wages, partial indexation of prices and wages, investments adjustment costs, habit persistence and variable capacity utilization. The endogenous variables in the model, expressed as log-deviations from the steady state, are output (y_t), consumption (c_t), investment (i_t), utilized and installed capital ($k_t^s,\ k_t$), capacity utilization (z_t), rental rate of capital (r_t^k), Tobin's q (q_t), price and wage mark-up ($\mu_t^p,\ \mu_t^w$), inflation rate(π_t), real wage (w_t), total hours worked (l_t), and nominal interest rate (r_t). The log-linearized equilibrium conditions for these variables are presented in Table A.1 in the Appendix. The last equation in the table gives the policy rule followed by the central bank, which sets the nominal interest rate in response to inflation and the deviation of output from its potential level. To determine potential output, defined as the level of output that would prevail in the absence of the price and wage mark-up shocks, the set of equations in Table A.1 is extended with their flexible price and wage version (see Table A.2 in the Appendix). The model has seven exogenous shocks. Five of them - total factor productivity, investment-specific technology, government purchases, risk premium, and monetary policy - follow AR(1) processes. The remaining two shocks - wage and price mark-up - follow ARMA$(1,1)$ processes. The model is estimated using data on seven variables: output growth, consumption growth, investment growth, real wage growth, inflation, hours worked and the nominal interest rate. Thus, the vector of observables is

$$\boldsymbol{x}_t = [y_t - y_{t-1}, c_t - c_{t-1}, i_t - i_{t-1}, w_t - w_{t-1}, \pi_t, l_t, r_t,]' \tag{4.1}$$

and the constant term in the measurement equation (2.2) is given by:

$$\boldsymbol{s}(\boldsymbol{\theta}) = \left[\bar{\gamma}, \bar{\gamma}, \bar{\gamma}, \bar{\gamma}, \bar{\pi}, \bar{l}, \bar{r}\right]' \tag{4.2}$$

where $\bar{\gamma}$ is the growth rate of output, consumption, investment and wages, $\bar{\pi}$ is the steady state rate of inflation, $\bar{l}$ is the steady state level of hours worked and $\bar{r}$ is the steady state nominal interest rate. Since there is no measurement error, the last term in (2.2) is omitted.

The deep parameters of the model are collected in a 41-dimensional vector $\boldsymbol{\theta}$ given by:[8]

$$\begin{aligned}\boldsymbol{\theta} &= [\delta, \lambda_w, g_y, \varepsilon_p, \varepsilon_w, \rho_{ga}, \beta, \mu_w, \mu_p, \alpha, \psi, \varphi, \sigma_c, \lambda, \Phi, \iota_w, \xi_w, \iota_p, \xi_p, \sigma_l, \\ &\quad r_\pi, r_{\triangle y}, r_y, \rho, \rho_a, \rho_b, \rho_g, \rho_I, \rho_r, \rho_p, \rho_w, \gamma, \sigma_a, \sigma_b, \sigma_g, \sigma_I, \sigma_r, \sigma_p, \sigma_w, \bar{\pi}, \bar{l}]' \end{aligned} \tag{4.3}$$

The definitions of the parameters are shown in Table A.3 in the Appendix.

4.2 Identification Analysis

The identifiability of the parameters in the SW07 model was studied in Iskrev (2010). There, it was found that 37 of the 41 parameters in (4.3) are locally identified. The remaining four parameters - ξ_w, ξ_p, ϵ_w andϵ_p, are not separately identifiable in the sense that, in the linearized model, ξ_w cannot be distinguished from ϵ_w, and ξ_p cannot be distinguished from ϵ_p. As in SW2007, in what follows I assume that ϵ_w and ϵ_p, as well as λ_w, δ and g_y, are known. This leaves 36 parameters whose identification will be analyzed. Details on the identification of all 39 identifiable parameters are presented in the Appendix.

4.2.1 Identification strength

The strength of identification of the free parameters in $\boldsymbol{\theta}$ is measured using the expected information matrix evaluated at the posterior mean reported in SW07 (see Table A.3) for $T = 156$, the sample size in SW07. The results are presented in panel A of Table 1, which has three columns. The first column, labeled "CRLB", shows the values of the Cramér-Rao lower bounds. The other two columns, labeled "Lb." and "Ub.", show the lower and upper bounds of one-standard deviation intervals around the posterior mean.

An examination of the values in the table shows that, among the structural parameters, as relatively weakly identified stand out the steady states of hours worked and inflation ($\bar{l}$ and $\bar{\pi}$), the discount factor (β), the elasticity of labor supply (σ_l), the price and wage indexation coefficients (ι_p and ι_w), the response to output gap in the monetary policy rule (r_y), and the investment adjustment cost parameter (φ). Among the structural shock parameters, the worst identified are the persistence coefficients of the monetary policy shock (ρ_r) and the risk premium shock (ρ_b). The best identified structural parameters are the steady state growth rate (γ), the interest rate smoothing coefficient (ρ), and the fixed cost in production parameter (Φ). Among the shock parameters the best identified are the government spending shock (ρ_g and σ_g), the productivity shock (ρ_a and σ_a), the persistence coefficients of the wage and price

[8]Note that, by definition, $\bar{\gamma} = 100(\gamma - 1)$, and $\bar{r}$ is determined from the values of β, σ_c, γ and $\bar{\pi}$ from $\bar{r} = 100(\frac{\bar{\pi}\gamma^{\sigma_c}}{\beta} - 1)$.

mark-up shocks (ρ_w and ρ_p), and the standard deviation of the monetary policy shock (σ_r).

Table 1: Identification strength at the posterior mean

param.	A. Cramér-Rao			B. Monte Carlo			C. Posterior		
	CRLB	Lb.	Ub	Std.	Lb.	Ub.	Std.	Lb.	Ub.
φ	1.878	3.866	7.622	1.728	4.016	7.472	1.029	4.715	6.773
σ_c	0.177	1.203	1.558	0.174	1.206	1.554	0.131	1.249	1.511
λ	0.059	0.655	0.773	0.061	0.653	0.775	0.042	0.672	0.755
ξ_w	0.078	0.623	0.778	0.075	0.626	0.776	0.071	0.630	0.771
σ_l	0.986	0.850	2.823	1.043	0.793	2.880	0.619	1.218	2.455
ξ_p	0.057	0.593	0.707	0.061	0.589	0.711	0.058	0.592	0.709
ι_w	0.206	0.383	0.795	0.186	0.403	0.775	0.133	0.456	0.722
ι_p	0.131	0.113	0.375	0.121	0.122	0.365	0.092	0.152	0.336
ψ	0.144	0.403	0.690	0.153	0.394	0.699	0.115	0.431	0.662
Φ	0.117	1.487	1.722	0.129	1.476	1.733	0.078	1.527	1.682
r_π	0.386	1.659	2.431	0.391	1.655	2.436	0.181	1.864	2.227
ρ	0.041	0.767	0.849	0.038	0.770	0.846	0.024	0.784	0.833
r_y	0.038	0.050	0.125	0.040	0.048	0.127	0.022	0.065	0.110
$r_{\triangle y}$	0.043	0.181	0.266	0.045	0.179	0.268	0.027	0.196	0.251
$\bar{\pi}$	0.227	0.558	1.012	0.199	0.586	0.984	0.098	0.688	0.883
$100\left(\beta^{-1}-1\right)$	0.125	0.041	0.291	0.114	0.052	0.280	0.060	0.106	0.227
$\bar{l}$	1.486	-0.945	2.028	1.212	-0.670	1.753	0.605	-0.064	1.147
γ	0.010	0.421	0.441	0.016	0.415	0.447	0.014	0.417	0.445
α	0.019	0.172	0.209	0.018	0.173	0.209	0.018	0.173	0.208
ρ_a	0.014	0.944	0.971	0.027	0.931	0.984	0.010	0.948	0.968
ρ_b	0.088	0.128	0.305	0.089	0.128	0.306	0.084	0.133	0.301
ρ_g	0.010	0.966	0.987	0.022	0.955	0.998	0.008	0.968	0.985
ρ_I	0.065	0.646	0.776	0.064	0.646	0.775	0.059	0.652	0.770
ρ_r	0.092	0.059	0.244	0.088	0.063	0.240	0.065	0.086	0.217
ρ_p	0.058	0.833	0.950	0.078	0.813	0.970	0.047	0.845	0.938
ρ_w	0.014	0.954	0.983	0.040	0.928	1.008	0.013	0.955	0.981
ρ_{ga}	0.100	0.422	0.621	0.102	0.419	0.623	0.089	0.432	0.610
μ_p	0.148	0.551	0.847	0.160	0.539	0.858	0.087	0.612	0.786
μ_w	0.059	0.782	0.901	0.076	0.766	0.917	0.051	0.790	0.893
σ_a	0.032	0.427	0.492	0.033	0.426	0.493	0.028	0.432	0.487
σ_b	0.026	0.214	0.267	0.025	0.215	0.266	0.023	0.217	0.264
σ_g	0.033	0.496	0.562	0.032	0.497	0.561	0.030	0.499	0.559
σ_I	0.049	0.404	0.502	0.052	0.401	0.505	0.048	0.405	0.502
σ_r	0.016	0.230	0.261	0.016	0.229	0.261	0.015	0.231	0.260
σ_p	0.022	0.118	0.162	0.021	0.119	0.161	0.017	0.123	0.157
σ_w	0.027	0.217	0.272	0.028	0.216	0.272	0.022	0.222	0.266

Note: CRLB in panel A shows the values of the Cramér-Rao lower bounds on the standard deviations of unbiased estimators. In panel B it is the standard deviation of the ML estimator obtained using Monte Carlo simulations with 1000 samples. In panel C it is the standard deviation of the posterior distribution. Lb. and Ub are the endpoints of one Std. intervals around the posterior mean.

The values in panel A of Table 1 are theoretical bounds on the sampling uncertainty and one-standard deviation intervals. In principle, the actual standard deviations and intervals could be larger. However, even in frequentist estimation of structural models, there are prior restrictions on the values of the parameters, which are ignored in the calculations of the information matrix and the corresponding Cramér-Rao (CR) lower bounds.[9] Consequently, the theoretical bounds may be greater than the actual values. To get a sense of how accurate

[9]In addition to the constraints implied by the economic meaning of some preference and technology parameters, there are also implicit restrictions on the values of some parameters implied by the assumption that the model has an unique solution.

the information matrix approach is in predicting the actual sampling uncertainty, I conduct a Monte Carlo (MC) simulation study where the free parameters in the model are estimated by MLE on 1000 samples with length $T = 156$ generated by the SW07 model with the values shown in Table A.3. The MC estimates of the standard deviations and the corresponding endpoints of the one-standard deviation intervals are presented in panel B of Table 1. Comparing the standard deviations with the theoretical lower bounds shows that the two are very similar for most parameter and, with a few exceptions, the CR values are smaller than the MC standard deviations. In the case of α, β and ρ, the exceptions can be explained with the *a priori* restrictions on these parameters being theoretically bounded between 0 and 1. In the case of $\bar{l}$ and $\bar{\pi}$, both of which are quite poorly identified, the MC standard deviations are smaller partly because of the restrictions on other parameters related to the steady state, particularly β, and partly because of the bounds imposed on these parameters in the simulations.[10] For several parameters, namely the autoregressive coefficients ρ_a, ρ_g and ρ_w, the MC standard deviations are significantly larger, by a factor of 2, than the CR bounds. This result can be explained by the fact that the true values of these parameters are close to the upper bound of 1, which introduces negative skewness in their MC distributions. As a result, the actual variances are larger than the ones implied by the local shapes of the distributions around the modes. If we impose symmetry and discard the values below the respective lower bounds (0.92 for ρ_a, 0.95 for ρ_g and 0.94 for ρ_w), the MC variances for the three parameters become very close to their theoretical lower bounds. These three parameters, together with γ, ρ_p and μ_w, for which the MC standard deviations exceed the CR bounds by between 25% and 50%, and $\bar{\pi}$ and $\bar{l}$, for which the CR bounds a larger by 16% and 24%, respectively, are the ones with the largest discrepancies between the MC standard deviations and the CR bounds. For the remaining 28 parameters, the difference between the MC standard deviations and the CR bounds is on average less than 1% and does not exceed 10% in either direction. The conclusion therefore is that, by and large, the measure of identification strength based on the expected information matrix provides an accurate indication of the actual sampling uncertainty of the ML estimator.

The last panel C of Table 1 shows the standard deviations of the posterior distribution of the parameters and the corresponding one-standard deviation intervals around the posterior mean. Although conceptually very different, comparing the Bayesian and frequentist intervals gives some idea about the contribution of the prior information in the estimation of the parameters. The Bayesian standard deviations are always smaller than the MC ones, on average by 53%. The largest differences are with respect to ρ_a, ρ_g and ρ_w - the same three parameters for which the MC standard deviations are largest relative to the CR lower bounds.

[10]The results in Table 1 were obtained assuming that both $\bar{l}$ and $\bar{\pi}$ are restricted between -20 and 20.

Among the structural parameters, the prior information plays a relatively large role with respect to r_π, μ_p, β and r_y, for which the differences between the Bayesian and MC standard deviations exceed 90%. The posterior and MC standard deviations are closes for ξ_p, α and ξ_w, among the structural parameters, and σ_I, σ_b, σ_g and σ_r, among the shock parameters. For these parameters, the MC standard deviations are at most 10% larger than the Bayesian ones.

Measuring the strength of identification on the basis of either the posterior or the MC standard deviations results in almost identical rankings of the parameters, in terms of their relative strength of identification, as when the CR bounds are used. This is not surprising given how similar the estimated Bayesian and MC standard deviations are to the theoretical lower bounds. At the same time, the use of prior information considerably improves the strength of identification of most parameters, and particularly that of $\bar{l}$, $\bar{\pi}$, r_π, β, and φ, whose posterior standard deviations are much smaller than MC standard deviations or the CR bounds.

As discussed in Section 3.2, the CR bound for a parameter θ_i is a product of two terms, the first of which depends on the sensitivity of the log-likelihood with respect to θ_i, and the second is related to the collinearity between the derivative of the log-likelihood with respect to θ_i and the derivatives with respect to the other free parameters. The sensitivity and collinearity factors in the decomposition are shown in Table 2 under the labels "sens." and "coll.", respectively, alongside the values of the CR bounds. To interpret the numbers, it helps to reiterate that the sensitivity factor for a parameter shows the value of the conditional CR bound, i.e. the bound when all other parameters are known and there is no collinearity. The collinearity factor shows how much larger is the bound when the other parameters are unknown. The results indicate that both factors could play important roles in determining the strength of identification. Parameters such as ρ_g, ρ_a, ρ_w, γ, σ_g and σ_r are very well identified because both the sensitivity and collinearity factors are small, meaning that each one of these parameters affects the log-likelihood in a strong and distinct way.[11] The opposite is true for parameters such as ρ_b, r_w, ι_p and σ_l. In the case of β and especially $\bar{l}$, identification is weak mostly because of the relatively high sensitivity factors, meaning that these parameters have a very weak effect on the log-likelihood function.[12] Large collinearity values significantly worsen the identification of parameters like μ_p, ξ_w, ρ_p, μ_w, and r_π. However, these effects are to some extent offset by small values of the sensitivity component which renders the overall strength of identification of these parameters relatively strong. In other words, these

[11]Note that, unlike the collinearity measure which is independent of the scale of the parameters, the sensitivity measure should be compared relative to the parameter values.

[12]Note that $\bar{l}$ affects the log-likelihood only through the mean of hours worked.

Table 2: IM decomposition the posterior mean

	CRLB	sens.	coll.	$\boldsymbol{\varrho}_i$	$\boldsymbol{\varrho}_{i(1)}$	$\boldsymbol{\varrho}_{i(2)}$	$\boldsymbol{\varrho}_{i(3)}$	$\boldsymbol{\varrho}_{i(4)}$
φ	1.878	0.964	1.947	0.858	0.718 (ρ_I)	0.824 (λ,ρ_I)	0.833 (λ,Φ,ρ_I)	0.836 (λ,Φ,ρ_I,σ_p)
σ_c	0.177	0.062	2.838	0.936	0.735 (λ)	0.777 (β,λ)	0.812 (β,λ,r_y)	0.836 (β,λ,σ_l,ρ_w)
λ	0.059	0.020	2.975	0.942	0.735 (σ_c)	0.801 (σ_c,$r_{\triangle y}$)	0.826 (σ_c,$r_{\triangle y}$,ρ_b)	0.856 (σ_c,ξ_w,σ_l,$r_{\triangle y}$)
ξ_w	0.078	0.017	4.550	0.976	0.906 (μ_w)	0.952 (μ_w,σ_l)	0.957 (μ_w,σ_l,ρ)	0.961 (μ_w,σ_l,ρ,ρ_w)
σ_l	0.986	0.411	2.402	0.909	0.774 (ξ_w)	0.813 (μ_w,ξ_w)	0.821 (μ_w,λ,ξ_w)	0.847 (μ_w,σ_c,λ,ξ_w)
ξ_p	0.057	0.023	2.525	0.918	0.796 (ρ_p)	0.828 (ξ_w,ρ_p)	0.853 (ξ_w,ρ_p,σ_w)	0.872 (ξ_w,r_π,ρ_p,σ_w)
ι_w	0.206	0.144	1.434	0.717	0.440 (σ_w)	0.573 (σ_p,σ_w)	0.653 (ξ_w,σ_p,σ_w)	0.687 (ξ_w,ρ_p,σ_p,σ_w)
ι_p	0.131	0.058	2.272	0.898	0.813 (μ_p)	0.851 (μ_p,σ_p)	0.867 (μ_p,ρ_p,σ_p)	0.881 (μ_w,μ_p,ρ_p,σ_p)
ψ	0.144	0.099	1.450	0.724	0.310 (α)	0.423 (α,$r_{\triangle y}$)	0.474 (α,$r_{\triangle y}$,r_y)	0.505 (α,λ,$r_{\triangle y}$,r_y)
Φ	0.117	0.064	1.840	0.839	0.461 (ξ_p)	0.559 (ξ_p,σ_a)	0.631 (α,ξ_p,σ_a)	0.668 (ψ,ξ_p,σ_a,σ_g)
r_π	0.386	0.121	3.185	0.949	0.566 (ρ)	0.866 (r_y,ρ)	0.910 (σ_c,r_y,ρ)	0.922 (σ_c,r_y,ρ,ρ_I)
ρ	0.041	0.016	2.621	0.924	0.566 (r_π)	0.825 (r_π,r_y)	0.868 (σ_c,r_π,r_y)	0.895 (σ_c,r_π,r_y,ρ_r)
r_y	0.038	0.014	2.732	0.931	0.508 (r_π)	0.807 (r_π,ρ)	0.882 (σ_c,r_π,ρ)	0.897 (σ_c,r_π,ρ,ρ_I)
$r_{\triangle y}$	0.043	0.022	1.965	0.861	0.593 (λ)	0.648 (ψ,λ)	0.691 (ψ,λ,σ_r)	0.737 (λ,r_π,ρ_r,σ_r)
$\bar{\pi}$	0.227	0.130	1.745	0.820	0.811 ($\bar{l}$)	0.817 ($\bar{l}$,β)	0.819 ($\bar{l}$,β,γ)	0.819 ($\bar{l}$,β,α,γ)
$100\left(\beta^{-1}-1\right)$	0.125	0.088	1.427	0.714	0.364 ($\bar{\pi}$)	0.436 ($\bar{\pi}$,α)	0.501 ($\bar{\pi}$,α,σ_c)	0.579 ($\bar{\pi}$,α,σ_c,λ)
$\bar{l}$	1.486	0.859	1.730	0.816	0.811 ($\bar{\pi}$)	0.815 ($\bar{\pi}$,γ)	0.816 ($\bar{\pi}$,β,γ)	0.816 ($\bar{\pi}$,β,α,γ)
γ	0.010	0.010	1.016	0.175	0.111 ($\bar{l}$)	0.145 ($\bar{l}$,$\bar{\pi}$)	0.173 ($\bar{l}$,$\bar{\pi}$,β)	0.175 ($\bar{l}$,$\bar{\pi}$,β,α)
α	0.019	0.013	1.399	0.699	0.310 (ψ)	0.422 (β,σ_c)	0.492 (β,ψ,σ_c)	0.542 (β,σ_c,ρ_a,ρ_g)
ρ_a	0.014	0.010	1.360	0.678	0.355 (ρ_g)	0.429 (σ_c,ρ_g)	0.503 (α,σ_c,ρ_g)	0.527 (α,ψ,σ_c,ρ_g)
ρ_b	0.088	0.042	2.115	0.881	0.827 (σ_b)	0.869 (λ,σ_b)	0.871 (σ_c,λ,σ_b)	0.872 (σ_c,λ,ρ,σ_b)
ρ_g	0.010	0.008	1.260	0.608	0.355 (ρ_a)	0.410 (σ_c,ρ_a)	0.482 (α,σ_c,ρ_a)	0.508 (α,ψ,σ_c,ρ_a)
ρ_I	0.065	0.028	2.301	0.901	0.840 (σ_I)	0.877 (φ,σ_I)	0.884 (φ,$r_{\triangle y}$,σ_I)	0.885 (φ,σ_c,λ,σ_I)
ρ_r	0.092	0.072	1.282	0.625	0.460 (ρ)	0.557 ($r_{\triangle y}$,ρ)	0.574 ($r_{\triangle y}$,ρ,σ_r)	0.584 ($r_{\triangle y}$,ρ,ρ_b,σ_r)
ρ_p	0.058	0.013	4.439	0.974	0.963 (μ_p)	0.968 (μ_p,ξ_p)	0.969 (μ_p,ι_p,ξ_p)	0.973 (μ_p,ι_p,ξ_p,σ_p)
ρ_w	0.014	0.008	1.890	0.848	0.761 (ξ_w)	0.796 (ξ_w,r_y)	0.805 (ψ,ξ_w,r_y)	0.812 (μ_w,ξ_w,r_y,σ_w)
ρ_{ga}	0.100	0.092	1.078	0.373	0.233 (Φ)	0.263 (Φ,ξ_p)	0.281 (Φ,ξ_p,σ_a)	0.292 (Φ,ξ_p,σ_a,σ_g)
μ_p	0.148	0.023	6.436	0.988	0.963 (ρ_p)	0.976 (ρ_p,σ_p)	0.986 (ι_p,ρ_p,σ_p)	0.987 (μ_w,ι_p,ρ_p,σ_p)
μ_w	0.059	0.016	3.698	0.963	0.906 (ξ_w)	0.935 (ξ_w,σ_w)	0.942 (ξ_w,r_π,σ_w)	0.947 (ξ_w,σ_l,r_π,σ_w)
σ_a	0.032	0.026	1.235	0.587	0.316 (Φ)	0.377 (ψ,Φ)	0.433 (ψ,Φ,ξ_p)	0.459 (α,ψ,Φ,ξ_p)
σ_b	0.026	0.014	1.920	0.854	0.827 (ρ_b)	0.836 (λ,ρ_b)	0.842 (σ_c,λ,ρ_b)	0.843 (β,σ_c,λ,ρ_b)
σ_g	0.033	0.030	1.092	0.401	0.215 (Φ)	0.262 (ψ,Φ)	0.300 (ψ,Φ,ξ_p)	0.333 (ψ,Φ,ξ_p,σ_a)
σ_I	0.049	0.026	1.915	0.853	0.840 (ρ_I)	0.841 (σ_l,ρ_I)	0.842 (σ_l,ρ_I,ρ_w)	0.843 (φ,λ,σ_l,ρ_I)
σ_r	0.016	0.014	1.133	0.470	0.294 ($r_{\triangle y}$)	0.344 (r_π,$r_{\triangle y}$)	0.371 (λ,r_π,$r_{\triangle y}$)	0.403 (r_π,$r_{\triangle y}$,r_y,ρ_r)
σ_p	0.022	0.008	2.807	0.934	0.880 (μ_p)	0.904 (μ_p,ι_p)	0.918 (μ_p,ι_w,ι_p)	0.925 (μ_p,ι_w,ι_p,ρ_p)
σ_w	0.027	0.014	1.992	0.865	0.725 (μ_w)	0.803 (μ_w,ι_w)	0.819 (μ_w,ι_w,ξ_p)	0.827 (μ_w,ι_w,ξ_p,ρ_w)

Note: sens. and coll. denote the sensitivity and collinearity components of CRLB shown in the first column (see equation (3.6)). $\boldsymbol{\varrho}_i$ is the multiple correlation between $\partial\ell_T(\boldsymbol{\theta})/\partial\theta_i$ and $\partial\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_{-i}$. $\boldsymbol{\varrho}_{i(n)}$ is the largest among all multiple correlation coefficients between $\partial\ell_T(\boldsymbol{\theta})/\partial\theta_i$ and $\partial\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_{-i}(n)$ for all possible combinations of n parameters from $\boldsymbol{\theta}_{-i}$. The selected parameters are shown in parentheses.

are parameters whose effects on the log-likelihood are quite strong but at the same time are not very distinct from the effects of other free parameters. For instance, in the case of μ_p, which is the MA coefficient of the price mark-up shock, the collinearity factor is around 6.4 which corresponds to a correlation coefficient $\boldsymbol{\varrho}_i = 0.988$, as can be seen from the fifth column in the table. In addition to the overall collinearity measure, one could compute correlation coefficients with respect to smaller sets of parameters and thus determine which ones among the remaining 35 free parameters most closely match the effect of μ_p on the log-likelihood. The largest correlation coefficients for groups of one to four parameters are shown in Table 2 under the labels "$\boldsymbol{\varrho}_{i(n)}$", for $1 \leq n \leq 4$. In the case of μ_p, the largest pairwise correlation coefficient is .963 with respect to the AR coefficient of the price mark-up shock ρ_p. Larger groups of parameters most collinear with μ_p include other price stickiness related parameters, namely σ_p and ι_p as well as the MA coefficient of the wage mark-up shock μ_w. Interestingly, even though the collinearity value for μ_w is also high, the largest pairwise correlation is with respect to the wage stickiness parameter, ξ_w, and not the AR coefficient of the wage mark-up shock. A larger group of functionally similar parameters to μ_w includes also the volatility of the wage mark-up shock σ_w, the response coefficient to inflation in the policy rule r_π and the elasticity of labor supply σ_l. Another result worth pointing out has to do with the question of whether it is possible to distinguish between monetary policy inertia and persistence in the monetary policy shock. This question can be answered by considering the correlation between the parameters ρ and ρ_r, which can be seen in Table 2 to be 0.46. Indeed, ρ is the parameter most similar to ρ_r in the way it affects the log-likelihood. However, the effects are far from identical and, as can also be seen in the table, the parameters most similar to ρ are other structural policy rule parameters, like r_π and r_y, as well as the elasticity of intertemporal substitution σ_c.[13]

Before concluding this section, it is worth briefly describing the main consequences of having three additional free parameters, namely λ_w, δ and g_y. As can be expected, the identification of most parameters is weaker due to the higher collinearity values compared to when the three parameters are fixed. By far, the worst affected is the price stickiness parameter ξ_w, whose CR bound is 2.43 times larger. The elasticity of intertemporal substitution σ_c, the discount factor β, the investment adjustment cost parameter φ, capacity utilization cost parameter ψ, the capital share α, fixed cost in production Φ, and the AR coefficient of the productivity shock ρ_a are all also strongly affected. For these parameters, the CR bounds are between 15 % and 66 % larger. Several parameters, such as γ, $\bar{l}$ and $\bar{\pi}$ are completely unaffected. More details are provided in Table A.6 in the Appendix, which, in addition to the

[13]Of course, this result applies only to the full information DSGE setting and says nothing about the identification of the parameters in a single equation setting.

CR bounds for all parameters, also shows the sensitivity and collinearity factors as well as the largest correlation coefficients for groups of one to four parameters. As can be seen there, the striking increase in the CR bound for ξ_w is due to the very large pairwise correlation of 0.95 of that parameter and λ_w. The two parameters are therefore difficult to distinguish on the basis of the log-likelihood, which would explain why λ_w was fixed in the first place.

4.2.2 Sources of identification

The ML estimator identifies the model parameters using information from first and second order moments of the data. With 156 observations on seven variables, there are 7630 such moments. This section determines which among them are the most informative moments for each parameter. As discussed in Section 3.3, the analysis is based on the weights assigned to moments in the first order conditions of the estimator.[14] The properties of the ML estimator imply that the weights are optimal in the sense that the information contained in the moments is used efficiently to obtain the most precise estimates possible.

The available first and second order moments are sorted according to the values of ω_{ij}, defined in (3.17), and the results for the first seven moments with largest weights for each parameter are shown in Table 3. In the case of the steady state parameters $\bar{\pi}$ and $\bar{l}$, the seven first order moments account for all of the weight.[15] A significant portion of the weight, 85% or more, is assigned to the first seven moments also for σ_a, ρ_{ga}, σ_g, γ, ψ, and α. However, for the parameters ρ_g, ρ_b, σ_c and ρ_I, only 50% or less of the weight is accounted for by the moments shown in the table. In any case, moments which are not shown receive very small weights individually and are therefore of no particular interest here.

The mean of inflation alone is very important for $\bar{\pi}$ and $\bar{l}$, where it accounts for 50% or more of the total weight. The mean of the interest rate is also very important for these parameters. These two moments are also very informative with respect to β, accounting together for about 50% of the total weight. Another parameter with a large amount of weight distributed among first order moments is γ, for which the mean of consumption is the most important, while the means of wages, interest rate, output and investment receive approximately the same weight. The only other parameter for which first order moments are relatively important is σ_c, with a 4% weight on the mean of the interest rate. For the remaining parameters, only second order moments receive significant weights. The variance and first order autocovariance of hours, in particular, are the most important moments for more than a third of the parameters. In the case of α, ψ, ρ_a, ρ_{ga}, σ_a, and σ_g, these two moments alone account for more than 50% of the

[14] More precisely, the weights are on the relative differences between empirical and theoretical moments. See equation (3.16).

[15] In the case of $\bar{\pi}$ the total weight exceeds 1 due to rounding.

total weight. The variance and first order autocovariance of inflation are also important for a large number of parameters, and especially for ξ_w, ρ_w, μ_w, σ_w, and σ_π. The variances and first order autocovariances of investment and the interest rate receive large weights in the first order conditions for the investment specific and interest rate shock parameters, respectively. Second order moments of the other three variables - output, consumption and wages - on their own are important in a few cases. The variance of wages is relatively important for σ_w, while the variances of output and consumption are important for σ_g and ρ_g, respectively. Individual cross moments receive large weights, 10% or more in several cases. The covariance of inflation and hours is very important for ι_w, Φ, ρ_y, ξ_p, ρ_p, μ_p, and σ_p. The covariance of the interest rate and inflation is important for ρ and r_π, while the covariance of investment and output is important for ρ_g.

It is interesting to compare the weights assigned by the ML estimator with the weights in the first order conditions of a GMM estimator with optimal weighting matrix (see Iskrev (2014) on how to compute the optimal weighting matrix for Gaussian DSGE models). Table A.4 in the Appendix shows the GMM weights when moments up to the first order autocovariances are used, i.e. 84 first and second order moments in total. Again, the first seven moments with largest weights for each parameter are shown. Comparing with the results in Table 3 shows that, with a few exceptions, the two estimators select as most important virtually the same sets of moments. There are relatively large differences in the cases of ι_w and γ, and smaller ones with respect to r_π, ι_p and ρ_I. In the case of ι_w, GMM assigns relatively more weight to the variance and autocovariance of h_t, instead of the same moments of π_t. With respect to γ, the mean of c_t is less important and that of w_t - more important, for the GMM than for the ML estimator. There are also some differences between the two estimators in the ordering and weights placed on different moments. However, considering the fact that MLE uses many more moments and the very different approaches for computing the weights, the consistency between the results in the two tables is remarkable.

Some comments regarding the importance of moments of hours worked are in order. Firstly, note that in the first order conditions for $\bar{l}$, the mean of h_t receives a much smaller weight than the means of π_t and r_t, even though $\bar{l}$ in the model is equal to $\mathrm{E}(h_t)$. This is due to two reasons. First, the model implies that the sample means of π_t and r_t are correlated with the estimate of $\mathrm{E}(h_t)$.[16] As a result, $\mathrm{E}(\pi_t)$ and $\mathrm{E}(r_t)$ receive non-zero weights in the first order conditions for $\bar{l}$, even though the derivatives of these two moments with respect to $\bar{l}$ are zero. Second, the model implies that the variance of the estimate of $\mathrm{E}(h_t)$ is much larger than the variances of the estimates of $\mathrm{E}(\pi_t)$ and $\mathrm{E}(r_t)$. Relative to the value of the mean, the variance

[16]This can be seen from matrix $\boldsymbol{\Sigma}$ in (3.23) which is the asymptotic covariance matrix of $\boldsymbol{\mu}$.

Table 3: The most informative moments

param.							
φ	$\text{var}(h_t, h_t)$ (0.18)	$\text{cov}(h_t, h_{t+1})$ (0.16)	$\text{var}(i_t, i_t)$ (0.06)	$\text{var}(r_t, r_t)$ (0.04)	$\text{var}(r_t, h_t)$ (0.04)	$\text{var}(i_t, c_t)$ (0.03)	$\text{var}(c_t, c_t)$ (0.03)
σ_c	$\text{var}(h_t, h_t)$ (0.12)	$\text{cov}(h_t, h_{t+1})$ (0.09)	$\text{var}(r_t, r_t)$ (0.08)	$\text{var}(r_t, h_t)$ (0.06)	$\text{var}(\pi_t, h_t)$ (0.06)	$E(r_t)$ (0.04)	$\text{cov}(r_t, r_{t+1})$ (0.04)
λ	$\text{var}(h_t, h_t)$ (0.16)	$\text{cov}(h_t, h_{t+1})$ (0.14)	$\text{var}(r_t, r_t)$ (0.09)	$\text{var}(r_t, h_t)$ (0.07)	$\text{cov}(h_t, r_{t+1})$ (0.06)	$\text{cov}(r_t, r_{t+1})$ (0.05)	$\text{var}(\pi_t, h_t)$ (0.03)
ξ_w	$\text{var}(\pi_t, \pi_t)$ (0.16)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.10)	$\text{var}(\pi_t, h_t)$ (0.09)	$\text{var}(r_t, \pi_t)$ (0.07)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(h_t, \pi_{t+1})$ (0.05)	$\text{cov}(r_t, \pi_{t+1})$ (0.03)
σ_l	$\text{var}(h_t, h_t)$ (0.24)	$\text{cov}(h_t, h_{t+1})$ (0.17)	$\text{var}(r_t, h_t)$ (0.09)	$\text{cov}(h_t, r_{t+1})$ (0.06)	$\text{var}(\pi_t, \pi_t)$ (0.04)	$\text{cov}(h_t, h_{t+2})$ (0.03)	$\text{cov}(\pi_t, h_t)$ (0.03)
ξ_p	$\text{var}(h_t, h_t)$ (0.25)	$\text{cov}(h_t, h_{t+1})$ (0.14)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{cov}(h_t, \pi_{t+1})$ (0.07)	$\text{cov}(h_t, h_{t+2})$ (0.06)	$\text{cov}(h_t, h_{t+3})$ (0.02)	$\text{var}(w_t, w_t)$ (0.02)
ι_w	$\text{cov}(\pi_t, \pi_{t+1})$ (0.14)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.07)	$\text{cov}(w_t, \pi_{t+1})$ (0.07)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(r_t, \pi_{t+1})$ (0.05)	$\text{var}(w_t, w_t)$ (0.04)
ι_p	$\text{var}(h_t, h_t)$ (0.18)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.11)	$\text{cov}(h_t, h_{t+1})$ (0.11)	$\text{var}(\pi_t, h_t)$ (0.08)	$\text{var}(\pi_t, \pi_t)$ (0.05)	$\text{cov}(h_t, h_{t+2})$ (0.04)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.03)
ψ	$\text{var}(h_t, h_t)$ (0.39)	$\text{cov}(h_t, h_{t+1})$ (0.37)	$\text{var}(r_t, h_t)$ (0.04)	$\text{cov}(h_t, r_{t+1})$ (0.03)	$\text{cov}(r_t, h_{t+1})$ (0.01)	$\text{var}(\pi_t, h_t)$ (0.01)	$\text{var}(i_t, i_t)$ (0.01)
Φ	$\text{var}(h_t, h_t)$ (0.15)	$\text{var}(\pi_t, h_t)$ (0.11)	$\text{cov}(h_t, h_{t+1})$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{var}(y_t, y_t)$ (0.04)	$\text{cov}(h_t, h_{t+2})$ (0.04)	$\text{var}(\pi_t, \pi_t)$ (0.03)
r_π	$\text{var}(\pi_t, \pi_t)$ (0.18)	$\text{var}(r_t, \pi_t)$ (0.10)	$\text{cov}(\pi_t, r_{t+1})$ (0.09)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.09)	$\text{var}(\pi_t, h_t)$ (0.08)	$\text{cov}(\pi_t, h_{t+1})$ (0.04)	$\text{var}(h_t, h_t)$ (0.04)
ρ	$\text{var}(r_t, \pi_t)$ (0.14)	$\text{var}(r_t, r_t)$ (0.13)	$\text{cov}(r_t, r_{t+1})$ (0.13)	$\text{cov}(\pi_t, r_{t+1})$ (0.08)	$\text{var}(\pi_t, \pi_t)$ (0.08)	$\text{var}(\pi_t, h_t)$ (0.07)	$\text{cov}(r_t, \pi_{t+1})$ (0.04)
r_y	$\text{var}(\pi_t, \pi_t)$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.08)	$\text{var}(r_t, \pi_t)$ (0.07)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{var}(h_t, h_t)$ (0.04)	$\text{var}(r_t, r_t)$ (0.03)
$r_{\triangle y}$	$\text{var}(h_t, h_t)$ (0.26)	$\text{cov}(h_t, h_{t+1})$ (0.19)	$\text{var}(r_t, h_t)$ (0.10)	$\text{cov}(h_t, r_{t+1})$ (0.08)	$\text{cov}(r_t, h_{t+1})$ (0.03)	$\text{cov}(h_t, h_{t+2})$ (0.03)	$\text{var}(r_t, r_t)$ (0.02)
$\bar{\pi}$	$E(\pi_t)$ (0.57)	$E(r_t)$ (0.28)	$E(c_t)$ (0.06)	$E(h_t)$ (0.03)	$E(y_t)$ (0.03)	$E(w_t)$ (0.02)	$E(i_t)$ (0.01)
β	$E(r_t)$ (0.30)	$E(\pi_t)$ (0.19)	$\text{var}(i_t, i_t)$ (0.06)	$\text{var}(i_t, y_t)$ (0.05)	$\text{var}(\pi_t, \pi_t)$ (0.04)	$E(c_t)$ (0.03)	$\text{var}(h_t, h_t)$ (0.03)
$\bar{l}$	$E(\pi_t)$ (0.50)	$E(r_t)$ (0.31)	$E(c_t)$ (0.06)	$E(h_t)$ (0.05)	$E(y_t)$ (0.04)	$E(w_t)$ (0.04)	$E(i_t)$ (0.02)
γ	$E(c_t)$ (0.25)	$E(w_t)$ (0.16)	$E(r_t)$ (0.14)	$E(y_t)$ (0.12)	$E(i_t)$ (0.10)	$E(\pi_t)$ (0.09)	$\text{var}(h_t, h_t)$ (0.02)
α	$\text{var}(h_t, h_t)$ (0.37)	$\text{cov}(h_t, h_{t+1})$ (0.36)	$\text{var}(i_t, i_t)$ (0.04)	$\text{var}(r_t, h_t)$ (0.03)	$\text{var}(i_t, y_t)$ (0.02)	$\text{cov}(h_t, r_{t+1})$ (0.02)	$\text{cov}(r_t, h_{t+1})$ (0.01)
ρ_a	$\text{cov}(h_t, h_{t+1})$ (0.27)	$\text{var}(h_t, h_t)$ (0.26)	$\text{var}(r_t, h_t)$ (0.03)	$\text{var}(\pi_t, h_t)$ (0.03)	$\text{var}(i_t, y_t)$ (0.02)	$\text{var}(c_t, y_t)$ (0.02)	$\text{cov}(h_t, \pi_{t+1})$ (0.01)
ρ_b	$\text{var}(r_t, r_t)$ (0.17)	$\text{var}(c_t, c_t)$ (0.08)	$\text{var}(r_t, c_t)$ (0.05)	$\text{cov}(r_t, r_{t+1})$ (0.04)	$\text{cov}(c_t, c_{t+1})$ (0.03)	$\text{cov}(i_t, i_{t+1})$ (0.03)	$\text{var}(i_t, i_t)$ (0.02)
ρ_g	$\text{var}(i_t, y_t)$ (0.10)	$\text{var}(i_t, i_t)$ (0.06)	$\text{var}(c_t, y_t)$ (0.06)	$\text{var}(c_t, c_t)$ (0.05)	$\text{var}(y_t, y_t)$ (0.03)	$\text{cov}(i_t, y_{t+1})$ (0.02)	$\text{cov}(c_t, y_{t+1})$ (0.02)
ρ_I	$\text{var}(i_t, i_t)$ (0.25)	$\text{cov}(i_t, i_{t+1})$ (0.08)	$\text{var}(i_t, c_t)$ (0.06)	$\text{var}(c_t, c_t)$ (0.03)	$\text{var}(\pi_t, h_t)$ (0.03)	$\text{var}(\pi_t, i_t)$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)
ρ_r	$\text{cov}(r_t, r_{t+1})$ (0.23)	$\text{var}(r_t, r_t)$ (0.16)	$\text{cov}(h_t, h_{t+1})$ (0.09)	$\text{cov}(r_t, r_{t+2})$ (0.08)	$\text{cov}(h_t, r_{t+1})$ (0.06)	$\text{var}(r_t, h_t)$ (0.05)	$\text{var}(h_t, h_t)$ (0.04)
ρ_p	$\text{var}(\pi_t, \pi_t)$ (0.13)	$\text{var}(h_t, h_t)$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.04)	$\text{cov}(h_t, h_{t+1})$ (0.04)	$\text{cov}(h_t, h_{t+2})$ (0.03)
ρ_w	$\text{var}(\pi_t, \pi_t)$ (0.18)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.12)	$\text{var}(\pi_t, h_t)$ (0.08)	$\text{var}(h_t, h_t)$ (0.06)	$\text{var}(r_t, \pi_t)$ (0.05)	$\text{cov}(h_t, \pi_{t+1})$ (0.04)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)
ρ_{ga}	$\text{var}(h_t, h_t)$ (0.46)	$\text{cov}(h_t, h_{t+1})$ (0.45)	$\text{var}(y_t, y_t)$ (0.03)	$\text{var}(c_t, y_t)$ (0.01)	$\text{var}(i_t, y_t)$ (0.01)	$\text{var}(h_t, c_t)$ (0.01)	$\text{var}(h_t, y_t)$ (0.01)
μ_p	$\text{var}(\pi_t, \pi_t)$ (0.14)	$\text{var}(h_t, h_t)$ (0.14)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.07)	$\text{cov}(h_t, \pi_{t+1})$ (0.04)	$\text{cov}(h_t, h_{t+2})$ (0.04)	$\text{cov}(h_t, h_{t+3})$ (0.03)
μ_w	$\text{var}(\pi_t, \pi_t)$ (0.20)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.08)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(h_t, \pi_{t+1})$ (0.05)	$\text{var}(r_t, \pi_t)$ (0.05)	$\text{var}(\pi_t, w_t)$ (0.03)
σ_a	$\text{var}(h_t, h_t)$ (0.49)	$\text{cov}(h_t, h_{t+1})$ (0.48)	$\text{var}(y_t, y_t)$ (0.01)	$\text{var}(h_t, y_t)$ (0.01)	$\text{cov}(y_t, h_{t+1})$ (0.01)	$\text{var}(h_t, w_t)$ (0.00)	$\text{cov}(w_t, h_{t+1})$ (0.00)
σ_b	$\text{var}(r_t, r_t)$ (0.18)	$\text{var}(c_t, c_t)$ (0.14)	$\text{cov}(r_t, r_{t+1})$ (0.10)	$\text{var}(r_t, c_t)$ (0.07)	$\text{cov}(c_t, r_{t+1})$ (0.03)	$\text{var}(r_t, h_t)$ (0.03)	$\text{cov}(i_t, i_{t+1})$ (0.02)
σ_g	$\text{var}(h_t, h_t)$ (0.35)	$\text{cov}(h_t, h_{t+1})$ (0.34)	$\text{var}(y_t, y_t)$ (0.06)	$\text{var}(c_t, y_t)$ (0.05)	$\text{var}(i_t, y_t)$ (0.05)	$\text{var}(c_t, c_t)$ (0.03)	$\text{var}(i_t, i_t)$ (0.02)
σ_I	$\text{var}(i_t, i_t)$ (0.26)	$\text{cov}(i_t, i_{t+1})$ (0.14)	$\text{cov}(h_t, h_{t+1})$ (0.08)	$\text{var}(h_t, h_t)$ (0.08)	$\text{var}(i_t, c_t)$ (0.04)	$\text{var}(r_t, i_t)$ (0.02)	$\text{var}(r_t, h_t)$ (0.02)
σ_r	$\text{cov}(r_t, r_{t+1})$ (0.20)	$\text{var}(r_t, r_t)$ (0.20)	$\text{var}(h_t, h_t)$ (0.09)	$\text{var}(r_t, h_t)$ (0.07)	$\text{cov}(h_t, h_{t+1})$ (0.07)	$\text{cov}(h_t, r_{t+1})$ (0.06)	$\text{var}(r_t, \pi_t)$ (0.06)
σ_p	$\text{var}(\pi_t, \pi_t)$ (0.17)	$\text{var}(h_t, h_t)$ (0.13)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.11)	$\text{cov}(h_t, h_{t+1})$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.07)	$\text{cov}(h_t, h_{t+2})$ (0.03)
σ_w	$\text{var}(\pi_t, \pi_t)$ (0.20)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.17)	$\text{var}(w_t, w_t)$ (0.11)	$\text{var}(r_t, \pi_t)$ (0.05)	$\text{var}(\pi_t, h_t)$ (0.05)	$\text{var}(\pi_t, w_t)$ (0.04)	$\text{cov}(h_t, \pi_{t+1})$ (0.04)

Note: Based on the weights assigned in the score.

for h_t is around 4, while for π_t and r_t, the ratios are around 0.07 and 0.04. Consequently, both the ML and GMM estimators place a much smaller weight on the deviation between sample and theoretical mean of h_t than on the deviations for the other two variables.

As with the sample mean, the model-implied variances of the sample second order moments of h_t are much larger than the variances of moments of the other variables. For example, relative to the true value of the variance of h_t, the variance of its sample estimate is around 2.[17] Among the other six variables, investment has the largest relative variance of around 0.17. Yet, second order moments of h_t, and in particular its variance, receive large weights in the first order conditions of the ML and GMM estimators. The precise reason for this is difficult to ascertain since, in addition to the variances, the first order condition weights depend in a complex way on the full correlation structure of the sample second order moments, as well as on the derivatives of the moments with respect to the structural parameters. The effect of the correlations among moments on the weights can be seen by comparing the results in Table 3 with those in Table A.5 in the Appendix, which shows the weights when the optimal weighting matrix is replaced by a diagonal matrix with the inverses of the variances of the moments on its main diagonal. This leads to some significant changes. In many cases, the moments with largest weights are different from the ones in Table 3. Note, however, that except the few parameters for which first order moments are very important, in all other cases, the largest weights are very small. This means that, now, many more moments, mostly autocovariances at different lags, receive nearly the same weights. From an estimation point of view, this is inefficient since such moments contribute relatively little independent information due to the strong collinearities among them. For $\bar{l}$ and the other steady state-related parameters, the concentration of weight occurs because very few moments, mostly means, are affected by these parameters. Naturally, for $\bar{l}$, all weight is on $\mathrm{E}(h_t)$, while for $\bar{\pi}$ the weight on $\mathrm{E}(r_t)$ is larger than the weight on $\mathrm{E}(\pi_t)$ because of the smaller variance of the sample mean of r_t.

Next, I examine the sources of identification from a frequency domain perspective. As explained earlier, I use the frequency domain approximation of the Fisher information matrix to compute Cramér-Rao bounds based on information from different subsets of frequencies.[18] The frequency band yielding the lowest bound for a given parameter is the most informative part of the spectrum for that parameter. Three non-overlapping frequency bands are considered: low frequencies (with period from 32 quarters to infinity), business cycle (BC) frequencies (with period from 6 to 32 quarters), and high frequencies (with period between

[17] To be clear, I am comparing $\mathrm{var}(\hat{m})/m$ where m is the moment, $\hat{m}$ is the sample estimate and var(.) is its asymptotic variance.

[18] To assess how good the approximation is, I compared the values of the frequency domain approximate Cramér-Rao bounds to their exact counterparts for different sample sizes. As can be seen from figure A.1 the frequency domain approximation is very accurate even for small sample sizes.

2 and 6 quarters). Before turning to the results, it should be pointed out that several parameters are not identified when information from frequency zero is not used. To address this, I fix three parameters - $\bar{\pi}$, $\bar{l}$ and γ.[19] Table 4 reports the Cramér-Rao bounds for each frequency band relative to the Cramér-Rao bounds from the full spectrum. On the basis of these ratios we can both determine the most informative parts of the spectrum and assess the loss of information when some frequencies are not used. Also, in addition to the bounds, the table shows the relative values of the two terms in the right hand side of the decomposition in equation (3.6), i.e the sensitivity and collinearity factors for the respective band of frequencies relative to the same factors with all frequencies. The decomposition helps to understand why a particular frequency band is most informative for a given parameter. Intuitively, one might expect that the most informative part of the spectrum for a parameter will be the one in which the parameter is very important, in the sense of having a very strong impact on the properties of the model in those frequencies. For example, parameters that determine the persistence of the shocks have stronger effect in the lower part of the spectrum when the process is very persistent. Thus, we may expect the low frequencies to be most informative about the autoregressive coefficients. The results in Table 4 show that this intuition is mostly correct. The autoregressive coefficients of the very persistent shocks (ρ_a, ρ_g, ρ_p and ρ_w, see Table A.3) have the strongest effects and are best identified from the low frequencies, while the much less persistent shocks parameter ρ_b is best identified from the high frequencies. Also, the coefficient of the not very persistent monetary policy shock, ρ_r, is much better identified in the high than in the low frequencies. The autoregressive coefficient of the investment specific shock, ρ_I, which falls somewhere in between in terms of persistence, has the strongest effect and is best identified in the BC frequencies. Overall, in most cases, the part of the spectrum with the smallest CRLBs is the one where the sensitivity component is also smallest.[20] In total, the low frequency are most informative for 11 parameters[21], the BC frequencies - for 24 parameters, and the high frequencies - for 1 parameter. Finally, the results reveal that, even when the most informative part of the spectrum is used, there is a significant loss of information compared to using the full spectrum. The smallest losses are between 30% and 40% in the cases of ρ_g, ρ_w and ρ_a. For many parameters the frequency band CRLBs are several times larger than the full spectrum ones.

[19] As shown in Iskrev (2010), $\gamma, \delta, \beta, \phi$, and λ are not simultaneously identifiable without the mean, i.e. frequency zero. Fixing one of them renders the four parameters identifiable. $\bar{\pi}$, $\bar{l}$ are clearly not identified without first order moments.

[20] In the case of the discount factor β, which, as one might expect, is best identified in the low frequencies, the relative sensitivity component in the low frequency band is shown to be 1 due to rounding. To be precise, it is equal to 1.034 meaning that β has some effect on the BC and high frequencies, but the effect is very weak.

[21] This includes the three fixed parameters which cannot be identified without frequency 0.

Table 4: CRLBs in the frequency domain

	low					BC					high				
φ	6.6	=	2.7	×	2.5	1.7	=	1.6	×	1.1	3.4	=	1.5	×	2.3
σ_c	2.1	=	1.4	×	1.4	1.7	=	1.6	×	1.1	7.1	=	3.0	×	2.4
λ	2.6	=	1.9	×	1.4	1.8	=	1.4	×	1.3	4.9	=	2.1	×	2.3
ξ_w	4.8	=	1.8	×	2.6	1.5	=	1.4	×	1.1	3.1	=	2.3	×	1.3
σ_l	2.5	=	1.9	×	1.3	1.6	=	1.4	×	1.2	3.8	=	2.3	×	1.7
ξ_p	9.3	=	2.0	×	4.7	1.8	=	1.4	×	1.3	5.2	=	2.1	×	2.4
ι_w	10.0	=	7.8	×	1.3	1.6	=	1.4	×	1.2	3.0	=	1.5	×	2.0
ι_p	23.8	=	3.4	×	6.9	1.5	=	1.2	×	1.3	3.5	=	2.3	×	1.6
ψ	2.0	=	1.5	×	1.3	2.0	=	1.7	×	1.2	4.9	=	2.2	×	2.2
Φ	2.1	=	2.2	×	0.9	1.7	=	1.4	×	1.2	5.8	=	1.9	×	3.0
r_π	2.1	=	1.5	×	1.4	1.6	=	1.5	×	1.0	5.4	=	2.6	×	2.1
ρ	3.1	=	2.0	×	1.6	1.5	=	1.4	×	1.1	4.4	=	2.2	×	2.0
r_y	2.0	=	1.2	×	1.7	2.0	=	2.2	×	0.9	8.7	=	4.1	×	2.1
$r_{\triangle y}$	4.4	=	2.6	×	1.7	1.6	=	1.4	×	1.1	2.6	=	1.7	×	1.5
$\bar{\pi}$			fixed					fixed					fixed		
β	1.5	=	1.0	×	1.5	13.1	=	4.4	×	3.0	43.8	=	9.0	×	4.8
$\bar{l}$			fixed					fixed					fixed		
γ			fixed					fixed					fixed		
α	2.8	=	1.8	×	1.5	4.4	=	1.4	×	3.2	15.7	=	2.5	×	6.3
ρ_a	1.4	=	1.2	×	1.2	3.0	=	2.3	×	1.3	13.5	=	4.4	×	3.0
ρ_b	77.0	=	2.5	×	31.3	1.9	=	1.2	×	1.6	2.9	=	2.7	×	1.1
ρ_g	1.3	=	1.1	×	1.2	3.0	=	2.6	×	1.2	9.8	=	4.4	×	2.2
ρ_I	11.7	=	1.8	×	6.5	1.6	=	1.4	×	1.2	9.6	=	2.3	×	4.1
ρ_r	18.2	=	2.8	×	6.5	1.9	=	1.5	×	1.3	2.4	=	1.5	×	1.6
ρ_p	2.5	=	1.6	×	1.6	3.9	=	1.5	×	2.6	12.3	=	2.6	×	4.8
ρ_w	1.4	=	1.3	×	1.1	2.8	=	1.9	×	1.5	9.8	=	2.9	×	3.4
ρ_{ga}	4.3	=	4.3	×	1.0	1.6	=	1.5	×	1.1	2.3	=	1.4	×	1.6
μ_p	24.6	=	2.3	×	10.5	2.4	=	1.2	×	1.9	6.4	=	2.4	×	2.6
μ_w	11.0	=	2.0	×	5.6	1.6	=	1.3	×	1.2	5.8	=	2.5	×	2.3
σ_a	4.5	=	4.4	×	1.0	1.9	=	1.5	×	1.2	5.9	=	1.4	×	4.2
σ_b	150.7	=	4.4	×	34.2	2.4	=	1.5	×	1.6	2.3	=	1.4	×	1.7
σ_g	4.2	=	4.2	×	1.0	1.6	=	1.5	×	1.1	1.9	=	1.4	×	1.3
σ_I	28.4	=	4.4	×	6.4	1.9	=	1.5	×	1.2	5.9	=	1.4	×	4.2
σ_r	30.5	=	4.2	×	7.2	1.7	=	1.5	×	1.2	2.6	=	1.4	×	1.9
σ_p	108.5	=	4.4	×	24.7	1.8	=	1.5	×	1.2	5.0	=	1.4	×	3.6
σ_w	34.1	=	4.2	×	8.2	1.9	=	1.5	×	1.3	2.7	=	1.4	×	1.9

Note: The CRLBs are decomposed as (see equation (3.6))

$$CRLB(\theta_i) = CRLB(\theta_i|\boldsymbol{\theta}_{-i}) \times (1/\sqrt{1-\boldsymbol{\varrho}_i^2})$$

where the first terms on the right-hand side is the value of the bound when all other parameters are known. The table shows the ratio of each term for the particular frequency band relative to its value using the full spectrum. The frequency bands are: $[0, \pi/16)$ (low), $[\pi/16, \pi/3]$ (business cycle - BC), $(\pi/3, \pi]$ (high), and $[0, \pi]$ (all). Note that $\bar{\pi}$, $\bar{l}$ and γ are identified only when frequency zero is included.

4.2.3 Extensions

In this section, I consider three extensions to the identification analysis of the SW07 model. In particular, I study the effects of changing parameter values, of using different observables, and of changing the sample size on the strength of identification. To make the comparison with the earlier results easier, I keep the values of all free parameters fixed. Thus, in the first exercise, I change the value of one of the calibrated parameters - the steady state wage mark-up λ_w. In the second exercise, I assume that data on inflation expectations $\mathrm{E}_t(\pi_{t+1})$ is available and sequentially replace each one of the original observables with it, keeping the

original parameter values. In the third, the parameter values and the set of observables are as before and the sample size T is varied. The purpose of these extensions is to show how identification is affected by the sample size, the choice of observable variables and the value at which λ_w is fixed. At the same time, it should be noted that if the model were to be estimated under these conditions, the point estimates would also change.

The effect of increasing and decreasing the value of λ_w, from 1.5 to 1.8 and 1.2 respectively, on the identification strength is measured with the change in the Cramér-Rao lower bounds of the free parameters. The results are presented in panel A of Table 5, which shows the bounds relative to their values at the original parameterization. As can be seen, the strength of identification of nearly all parameters is affected, although to different degrees and in different directions. The most affected parameters are $\bar{\pi}$ $\bar{l}$, whose CR bounds change by about 40% when λ_w is changed in either direction. ρ_w, ξ_w and σ_l, whose bounds change by about 30% when λ_w is decreased from 1.5 to 1.2, are also strongly affected. In all cases, decreasing the value of λ_w has the opposite effect on the CR bounds from that of increasing it. Relatively more parameters are better identified when λ_w is smaller and the effect on them is stronger than the effect on the parameters whose identification is weaker with the lower value of λ_w.

Panel B of Table 5 shows the effect on the CR bounds of replacing variables from the original set of observables with inflation expectations. Again, the ratios of the new to the original bounds are presented. Since $\bar{l}$ is not identified when h_t is not observed, that variable is always included. Several results are worth highlighting. First, neither one of the seven possible sets of observables (including the original one) dominates in the sense of giving the lowest CR bounds for all parameters. For 23 of the 36 parameters the best set of observables is the one where consumption is replaced with inflation expectations, i.e. observing $y_t, I_t, w_t, h_t, \pi_t, r_t$ and $\mathrm{E}_t(\pi_{t+1})$. That set of variables is also best in terms of overall strength of identification, as measured by the geometric average of the CR bounds. Second, the parameters whose identification improves the most from observing $\mathrm{E}_t(\pi_{t+1})$ instead of c_t are μ_p, ρ_p, ι_p, ρ_r, and ρ, whose CR bounds are 70% or more smaller, compared to their original values. However, for other parameters, such as σ_b, σ_c, ρ_{ga}, and α, the CR bounds increase as a result of replacing c_t with $\mathrm{E}_t(\pi_{t+1})$. Third, as might be expected, replacing π_t with $\mathrm{E}_t(\pi_{t+1})$ has a relatively small effect on the strength of identification of most parameters. Several of them, such as σ_p, ι_p, ι_w and σ_w are identified much worse, while others - ρ_p, r_π, r_y, $\bar{\pi}$, μ_w and ρ_w - are better identified when $\mathrm{E}_t(\pi_{t+1})$ is observed instead of π_t.

The last exercise considers the effect of changing the sample size T on the strength of identification. Naturally, more data means more information and therefore more precise

Table 5: Effect of changing the value of λ_w, the set of observables, and the sample size

	A. Value of λ_w		B. Variable replaced with E(π)						C. Rate of convergence
param.	$\lambda_w = 1.2$	$\lambda_w = 1.8$	r	π	w	I	c	y	
φ	1.08	0.94	0.96	0.99	0.96	1.67	0.98	0.94	$b = 0.50$
σ_c	0.89	1.07	1.08	1.00	0.99	0.88	1.25	0.98	b = 0.51
λ	1.05	0.92	1.04	1.00	0.98	0.80	1.01	0.98	$b = 0.51$
ξ_w	0.71	1.14	0.43	1.00	1.32	0.48	0.37	0.63	$b = 0.51$
σ_l	0.75	1.05	0.50	1.01	1.04	0.56	0.62	0.79	$b = 0.51$
ξ_p	1.00	0.99	0.44	1.02	1.18	0.43	0.43	0.67	$b = 0.51$
ι_w	0.96	1.00	0.44	1.33	4.27	0.47	0.32	0.64	$b = 0.50$
ι_p	1.15	0.89	0.23	1.54	0.78	0.26	0.16	0.33	$b = 0.50$
ψ	1.03	0.95	0.78	1.01	1.09	0.76	0.56	1.06	$b = 0.55$
Φ	1.14	0.88	0.56	0.99	1.11	0.58	0.50	1.19	$b = 0.51$
r_π	0.98	1.03	0.69	0.94	0.91	0.56	0.44	0.77	$b = 0.51$
ρ	0.96	1.02	0.74	1.04	0.94	0.45	0.30	0.69	$b = 0.51$
r_y	1.08	0.98	0.74	0.94	0.95	0.64	0.60	0.78	$b = 0.51$
$r_{\triangle y}$	0.97	0.99	1.71	1.00	0.98	0.87	0.64	0.89	$b = 0.50$
$\bar{\pi}$	0.59	1.44	0.77	0.94	0.78	0.77	0.77	0.77	$b = 0.44$
β	0.97	1.03	6.23	1.03	0.99	0.93	1.07	0.99	$b = 0.49$
$\bar{l}$	0.64	1.37	0.88	1.00	0.90	0.90	0.89	0.90	$b = 0.45$
γ	1.03	0.98	0.27	1.01	1.08	0.56	0.45	0.82	$b = 1.55$
α	1.03	0.96	1.98	1.00	1.02	3.33	1.17	1.23	$b = 0.50$
ρ_a	1.03	0.97	0.57	1.00	1.05	0.99	0.52	0.86	$b = 0.54$
ρ_b	1.02	0.98	0.93	1.00	1.00	0.87	0.77	0.89	$b = 0.50$
ρ_g	0.97	1.02	0.95	0.99	0.98	0.93	0.86	1.03	$b = 0.57$
ρ_I	1.01	0.99	0.53	0.96	0.90	0.64	0.44	0.78	$b = 0.50$
ρ_r	1.00	1.00	8.10	1.03	0.98	0.56	0.17	0.74	$b = 0.50$
ρ_p	1.02	0.98	0.11	0.88	1.05	0.11	0.11	0.22	$b = 0.52$
ρ_w	1.31	0.85	0.52	0.95	1.17	0.58	0.48	0.55	$b = 0.51$
ρ_{ga}	1.02	0.98	0.99	1.00	1.00	1.35	1.23	1.60	$b = 0.50$
μ_p	1.07	0.96	0.10	1.07	0.84	0.12	0.05	0.21	$b = 0.51$
μ_w	0.92	1.05	0.26	0.94	3.05	0.28	0.22	0.41	$b = 0.50$
σ_a	1.04	0.96	0.87	1.00	1.04	0.98	0.85	1.36	$b = 0.51$
σ_b	1.00	0.99	0.82	1.00	1.00	0.85	1.54	0.92	$b = 0.50$
σ_g	1.03	0.98	1.01	1.00	1.03	1.26	1.29	1.76	$b = 0.50$
σ_I	0.98	1.02	0.87	0.99	0.97	3.89	0.71	0.86	$b = 0.51$
σ_r	0.99	1.00	16.29	1.00	1.00	0.96	1.01	0.98	$b = 0.50$
σ_p	0.93	1.05	0.38	2.23	0.56	0.38	0.37	0.40	$b = 0.50$
σ_w	0.85	1.15	0.57	1.18	7.20	0.58	0.56	0.64	$b = 0.50$

Note: Panels A and B show the Cramér-Rao lower bounds after the value of λ_w or the set of observables is changed, relative to the original bounds. The original value of λ_w is 1.5. The original set of observables contains y, c, I, w, h, π and r. Panel C shows the values of b in the power function aT^{-b} which describes the behavior of the CR bounds as functions of the sample size T for $200 \leq T \leq 1000$.

estimates. Thus, the values of the CR bounds on the standard deviations should decrease as T increases. However, it may be interesting to know the size of the gains one may expect with respect to different parameters. Furthermore, we could measure the rates at which the bounds shrink as the sample size grows. It is standard in the econometric literature to define weak identification as having sample information about one or more parameters accumulating at rates slower than $\sqrt{T}$ (see e.g. Stock and Wright (2000)). Similarly, in the context of Bayesian estimation, Koop et al. (2013) deem as weakly identified the parameters whose posterior precision updates at rates slower than $\sqrt{T}$. To see the effect of increasing the sample size, I compute the CR bounds in 50 points for T between 200 and 1000. The results are shown in Figures A.2 and A.3 in the Appendix. The plots suggest that the behavior of the bounds may be well described by a power function of the form aT^{-b}, with $a > 0$ and $b > 0$.

A more formal regression analysis shows that this is indeed the case, and the power function performs very well both in terms of fitting and forecasting the observed patterns.[22] Panel C of Table 5 shows the values of the b coefficients for all parameters. With the exception of $\bar{l}$ and $\bar{\pi}$, whose convergence is somewhat slower, and γ, whose convergence rate is significantly faster, the CR bounds for all other parameters exhibit rates of convergence very close to $\sqrt{T}$. When the same exercises is repeated for smaller sample sizes, i.e. with T between 50 and 200 - the rates of convergence for $\bar{l}$ and $\bar{\pi}$ are significantly slower, with $b \approx .3$. For all other parameters, the rates of convergence change very little and in most cases increase when T is smaller. Thus, with the possible exceptions of $\bar{l}$ and $\bar{\pi}$, none of the other parameters would qualify as weakly identified in the sense of information accumulating at slower rate than $\sqrt{T}$.[23]

5 Concluding Remarks

There are two main reasons why we should care about identification in DSGE models. First, using such models for policy analysis hinges upon having reliably estimated parameters. Obtaining such estimates is impossible when identification fails or is very weak. Second, identification failures are often rooted in the underlying model and the economic theory that motivates it. Thus, detecting identification problems and investigating the causes leading to them may provide useful insight to researchers who are not interested in estimation.

This paper develops a new framework for analyzing parameter identification in linearized DSGE models. By following the steps and applying the tools described here, researchers can assess how well identified their model parameters are, as well as determine the causes for identification problems when they occur. Also, the consequences for identification from changing parameter values, the set of observables, and the sample size can be explored. An important advantage of the methodology is that it does not involve simulation or estimation. This makes it suitable for analysis of large and complicated models prior to their empirical evaluation.

Although this paper focused on the identification of parameters, it is straightforward to extend the analysis to any other model object that can be expressed as a function of the parameters. Such objects of interest include impulse response functions, moments of observed and unobserved variables, variance decompositions, and smoothed structural shocks.

One limitation of this paper's approach is that it cannot detect certain types of global

[22] To see how well the power function predicts the values of the CR bounds, I estimated the coefficients using the first 40 points and compared the predicted values with the last 10 points.

[23] Qu (2014) characterizes weak identification as having some eigenvalues of the normalized information matrix converging to zero as T increases. As shown in the Appendix, this condition does not hold in the case of the SW07 model, although three of the eigenvalues remain close to but strictly above zero as T increases.

identification problems. It is possible that some parameters are well identified locally, and yet are unidentifiable or poorly identified globally. Such identification failures are less common, but not impossible. Unfortunately, they are very difficult to discover in large and highly non-linear models such as those estimated in the DSGE literature.

Appendix

Evaluating (3.25)

To evaluate (3.25) we need to compute $\partial \boldsymbol{F}(\omega)/\partial\theta_k$. Note that the spectral density matrix for the model in (2.2)-(2.3) is

$$\boldsymbol{F}(\omega) = \boldsymbol{C}\boldsymbol{M}^{-}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{M}^{+}\boldsymbol{C}' \tag{A.1}$$

where $\boldsymbol{M}^{-} := (\boldsymbol{I}_m - \boldsymbol{A}\exp(-i\omega))^{-1}$ and $\boldsymbol{M}^{+} := (\boldsymbol{I}_m - \boldsymbol{A}\exp(i\omega))^{-1}$. Denoting the derivative of a matrix $\boldsymbol{X}$ w.r.t. θ_i by $\partial_i\boldsymbol{X}$, we have

$$\begin{aligned}\partial_i\boldsymbol{F}(\omega) = \partial_i\boldsymbol{C}\boldsymbol{M}^{-}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{M}^{+}\boldsymbol{C}' + \boldsymbol{C}\partial_i\boldsymbol{M}^{-}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{M}^{+}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{M}^{-}\partial_i\boldsymbol{B}\boldsymbol{B}'\boldsymbol{M}^{+}\boldsymbol{C}' + \\ \boldsymbol{C}\boldsymbol{M}^{-}\boldsymbol{B}\partial_i\boldsymbol{B}'\boldsymbol{M}^{+}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{M}^{-}\boldsymbol{B}\boldsymbol{B}'\partial_i\boldsymbol{M}^{+}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{M}^{-}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{M}^{+}\partial_i\boldsymbol{C}'\end{aligned}$$

Using $\partial_i\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}\partial_i\boldsymbol{X}\boldsymbol{X}^{-1}$, we have

$$\partial_i\boldsymbol{M}^{-} = -\boldsymbol{M}^{-}\left(\boldsymbol{I}_m - \partial_i\boldsymbol{A}\exp(-i\omega)\right)\boldsymbol{M}^{-} \tag{A.2}$$

and similarly for $\boldsymbol{M}^{+}$. Lastly, to complete the evaluation of $\partial_i\boldsymbol{F}(\omega)$ we need the derivatives of $\partial_i\boldsymbol{A}$, $\partial_i\boldsymbol{B}$, and $\partial_i\boldsymbol{C}$, which can be obtained as shown in Iskrev (2008)

Derivation of (3.11)

The innovation representation of the state space system (2.2)-(2.3) is

$$\begin{aligned}\hat{\boldsymbol{z}}_{t|t-1} &= \boldsymbol{A}\hat{\boldsymbol{z}}_{t-1|t-2} + \boldsymbol{K}_{t-1}\boldsymbol{e}_{t-1|t-2} &\text{(A.3)}\\ \boldsymbol{x}_t &= \boldsymbol{s} + \boldsymbol{C}\hat{\boldsymbol{z}}_{t|t-1} + \boldsymbol{e}_{t|t-1} &\text{(A.4)}\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{K}_t &= \boldsymbol{A}\boldsymbol{P}_{t|t-1}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}'\right)^{-1} &\text{(A.5)}\\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{A}\boldsymbol{P}_{t|t-1}\boldsymbol{A}' - \boldsymbol{K}_t\left(\boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}'\right)\boldsymbol{K}_t' + \boldsymbol{B}\boldsymbol{B}' &\text{(A.6)}\end{aligned}$$

Expanding the recursion in equation (A.3) to substitute for $\hat{\boldsymbol{z}}_{t|t-1}$ in (A.4), we have

$$\boldsymbol{x}_t - \boldsymbol{s} = \sum_{h=1}^{t-1}\boldsymbol{C}\boldsymbol{A}^h\boldsymbol{K}_{t-h}\boldsymbol{e}_{t-h|t-h-1} + \boldsymbol{C}\boldsymbol{A}^{t-1}\hat{\boldsymbol{z}}_{1|0} + \boldsymbol{e}_{t|t-1} \tag{A.7}$$

Using (A.7) and the initial condition $\hat{\boldsymbol{z}}_{1|0} = \boldsymbol{0}$, it follows that

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_1 - \boldsymbol{s} \\ \boldsymbol{x}_2 - \boldsymbol{s} \\ \boldsymbol{x}_3 - \boldsymbol{s} \\ \vdots \\ \boldsymbol{x}_T - \boldsymbol{s} \end{bmatrix}}_{\boldsymbol{X}_T} = \underbrace{\begin{bmatrix} \boldsymbol{I} & \mathbf{O} & \dots & \mathbf{O} \\ \boldsymbol{CAK}_1 & \boldsymbol{I} & \dots & \mathbf{O} \\ \boldsymbol{CA}^2\boldsymbol{K}_1 & \boldsymbol{CAK}_2 & \dots & \mathbf{O} \\ \vdots & \vdots & & \mathbf{O} \\ \boldsymbol{CA}^{T-1}\boldsymbol{K}_1 & \boldsymbol{CA}^{T-2}\boldsymbol{K}_2 & \dots & \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{L}} \underbrace{\begin{bmatrix} \boldsymbol{e}_{1|0} \\ \boldsymbol{e}_{2|1} \\ \boldsymbol{e}_{3|2} \\ \vdots \\ \boldsymbol{e}_{T|T-1} \end{bmatrix}}_{\boldsymbol{E}_T} \tag{A.8}$$

Table A.1: Log-linearized equations of the SW07 model (sticky-price-wage economy)

(1) $y_t = c_y c_t + i_y i_t + r^{kss} k_y z_t + \varepsilon_t^g$

(2) $c_t = \frac{\lambda/\gamma}{1+\lambda/\gamma} c_{t-1} + \frac{1}{1+\lambda/\gamma} \mathrm{E}_t\, c_{t+1} + \frac{w^{ss} l^{ss} (\sigma_c - 1)}{c^{ss} \sigma_c (1+\lambda/\gamma)} (l_t - \mathrm{E}_t\, l_{t+1})$
$- \frac{1-\lambda/\gamma}{(1+\lambda/\gamma)\sigma_c} (r_t - \mathrm{E}_t\, \pi_{t+1}) - \frac{1-\lambda/\gamma}{(1+\lambda/\gamma)\sigma_c} \varepsilon_t^b$

(3) $i_t = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} i_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} \mathrm{E}_t\, i_{t+1} + \frac{1}{\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})} q_t + \varepsilon_t^i$

(4) $q_t = \beta(1-\delta)\gamma^{-\sigma_c} \mathrm{E}_t\, q_{t+1} - r_t + \mathrm{E}_t\, \pi_{t+1} + (1 - \beta(1-\delta)\gamma^{-\sigma_c}) \mathrm{E}_t\, r_{t+1}^k - \varepsilon_t^b$

(5) $y_t = \phi_p(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a)$

(6) $k_t^s = k_{t-1} + z_t$

(7) $z_t = \frac{1-\psi}{\psi} r_t^k$

(8) $k_t = (1-\delta)/\gamma k_{t-1} + (1 - (1-\delta)/\gamma) i_t + (1 - (1-\delta)/\gamma)\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})\varepsilon_t^i$

(9) $\mu_t^p = \alpha(k_t^s - l_t) - w_t + \varepsilon_t^a$

(10) $\pi_t = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\iota_p\beta\gamma^{(1-\sigma_c)}} \mathrm{E}_t\, \pi_{t+1} + \frac{\iota_p}{1+\beta\gamma^{1-\sigma_c}\iota_p} \pi_{t-1} - \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_p)(1-\xi_p)}{(1+\iota_p\beta\gamma^{(1-\sigma_c)})(1+(\phi_p-1)\varepsilon_p)\xi_p} \mu_t^p + \varepsilon_t^p$

(11) $r_t^k = l_t + w_t - k_t$

(12) $\mu_t^w = w_t - \sigma_l l_t - \frac{1}{1-\lambda/\gamma}(c_t - \lambda/\gamma c_{t-1})$

(13) $w_t = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} (\mathrm{E}_t\, w_{t+1} + \mathrm{E}_t\, \pi_{t+1}) + \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} (w_{t-1} + \iota_w \pi_{t-1}) - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}} \pi_t$
$- \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{(1+\beta\gamma^{(1-\sigma_c)})(1+(\phi_w-1)\varepsilon_w)\xi_w} \mu_t^w + \varepsilon_t^w$

(14) $r_t = \rho r_{t-1} + (1-\rho)(r_\pi \pi_t + r_y(y_t - y_t^*)) + r_{\triangle y}((y_t - y_t^*) - (y_{t-1} - y_{t-1}^*)) + \varepsilon_t^r$

(15) $\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \eta_t^a$

(16) $\varepsilon_t^b = \rho_a \varepsilon_{t-1}^b + \eta_t^b$

(17) $\varepsilon_t^g = \rho_g \varepsilon_{t-1}^a + \rho_{ga} \eta_t^a + \eta_t^g$

(18) $\varepsilon_t^i = \rho_I \varepsilon_{t-1}^I + \eta_t^I$

(19) $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \eta_t^r$

(20) $\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \eta_t^p - \mu_p \eta_{t-1}^p$

(21) $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w$

Note: The model variables are: output (y_t), consumption (c_t), investment (i_t), utilized and installed capital ($k_t^s,\ k_t$), capacity utilization (z_t), rental rate of capital (r_t^k), Tobin's q (q_t), price and wage markup ($\mu_t^p,\ \mu_t^w$), inflation rate(π_t), real wage (w_t), total hours worked (l_t), and nominal interest rate (r_t). The shocks are: total factor productivity (ε_t^a), investment-specific technology (ε_t^i), government purchases (ε_t^g), risk premium (ε_t^b), monetary policy (ε_t^r), wage markup (ε_t^w) and price markup (ε_t^p).

Table A.2: Log-linearized equations of the SW07 model (flexible-price-wage economy)

(1) $y_t^* = c_y c_t^* + i_y i_t^* + r^{kss} k_y z_t^* + \varepsilon_t^g$

(2) $c_t^* = \frac{\lambda/\gamma}{1+\lambda/\gamma} c_{t-1}^* + \frac{1}{1+\lambda/\gamma} \mathrm{E}_t\, c_{t+1}^* + \frac{w^{ss} l^{ss} (\sigma_c - 1)}{c^{ss} \sigma_c (1+\lambda/\gamma)} (l_t^* - \mathrm{E}_t\, l_{t+1}^*)$
$- \frac{1-\lambda/\gamma}{(1+\lambda/\gamma)\sigma_c} r_t^* - \frac{1-\lambda/\gamma}{(1+\lambda\lambda/\gamma)\sigma_c} \varepsilon_t^b$

(3) $i_t^* = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} i_{t-1}^* + \frac{\beta\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} \mathrm{E}_t\, i_{t+1}^* + \frac{1}{\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})} q_t^* + \varepsilon_t^i$

(4) $q_t^* = \beta(1-\delta)\gamma^{-\sigma_c} \mathrm{E}_t\, q_{t+1}^* - r_t^* + (1 - \beta(1-\delta)\gamma^{-\sigma_c}) \mathrm{E}_t\, r_{t+1}^{k*} - \varepsilon_t^b$

(5) $y_t^* = \phi_p(\alpha k_t^{s*} + (1-\alpha) l_t^* + \varepsilon_t^a)$

(6) $k_t^{s*} = k_{t-1}^* + z_t^*$

(7) $z_t^* = \frac{1-\psi}{\psi} r_t^{k*}$

(8) $k_t^* = (1-\delta)/\gamma k_{t-1}^* + (1 - (1-\delta)/\gamma) i_t^* + (1 - (1-\delta)/\gamma)\varphi\gamma^2(1 + \beta\gamma^{(1-\sigma_c)})\varepsilon_t^i$

(9) $\mu_t^{p*} = \alpha(k_t^{s*} - l_t^*) - w_t^* + \varepsilon_t^a$

(10) $\mu_t^{p*} = 1$

(11) $r_t^{k*} = l_t^* + w_t^* - k_t^*$

(12) $\mu_t^{w*} = -\sigma_l l_t^* - \frac{1}{1-\lambda/\gamma}(c_t^* + \lambda/\gamma c_{t-1}^*)$

(13) $w_t^* = \mu_t^{w*}$

Note: The model variables are: output (y_t^*), consumption (c_t^*), investment (i_t^*), utilized and installed capital ($k_t^{s*},\ k_t^*$), capacity utilization (z_t^*), rental rate of capital (r_t^{k*}), Tobin's q (q_t^*), price and wage markup ($\mu_t^{p*},\ \mu_t^{w*}$), real wage (w_t^*), and total hours worked (l_t^*).

Table A.3: Parameters in SW07

parameter	interpretation	posterior mean
φ	investment adjustment cost	5.744
σ_c	elasticity of intertemporal substitution	1.380
λ	habit	0.714
ξ_w	wage stickiness	0.701
σ_l	elasticity of labor supply	1.837
ξ_p	price stickiness	0.650
ι_w	wage indexation	0.589
ι_p	price indexation	0.244
ψ	capacity utilization cost	0.546
Φ	fixed cost in production	1.604
r_π	monetary policy response to inflation	2.045
ρ	interest rate smoothing	0.808
r_y	monetary policy response to output gap	0.088
$r_{\triangle y}$	monetary policy response to change in output gap	0.224
$\bar{\pi}$	steady state inflation	0.785
$100\left(\beta^{-1}-1\right)$	discount factor	0.166
$\bar{l}$	steady state hours	0.542
γ	trend growth rate	0.431
α	capital share	0.191
ρ_a	AR productivity shock	0.958
ρ_b	AR risk premium shock	0.217
ρ_g	AR government spending shock	0.976
ρ_I	AR investment specific shock	0.711
ρ_r	AR monetary policy shock	0.151
ρ_p	AR price markup shock	0.891
ρ_w	AR wage markup shock	0.968
ρ_{ga}	productivity shock in G	0.521
μ_p	MA price markup	0.699
μ_w	MA wage markup	0.841
σ_a	standard deviation productivity shock	0.460
σ_b	standard deviation risk premium shock	0.240
σ_g	standard deviation government spending shock	0.529
σ_I	standard deviation investment specific shock	0.453
σ_r	standard deviation monetary policy shock	0.245
σ_p	standard deviation price markup shock	0.140
σ_w	standard deviation wage markup shock	0.244
$\delta^\dagger$	depreciation rate	0.025
${\lambda_w}^\dagger$	wage markup	1.500
${g_y}^\dagger$	steady state government spending-output ratio	0.180
${\varepsilon_p}^\dagger$	curvature of goods market aggregator	10.000
${\varepsilon_w}^\dagger$	curvature of labor market aggregator	10.000

† These parameters are assumed known in SW07.

Table A.4: Moments with largest weights in the GMM first order conditions

param.							
φ	$\text{var}(h_t, h_t)$ (0.18)	$\text{cov}(h_t, h_{t+1})$ (0.17)	$\text{var}(i_t, i_t)$ (0.06)	$\text{var}(r_t, r_t)$ (0.05)	$\text{cov}(r_t, r_{t+1})$ (0.05)	$\text{var}(r_t, h_t)$ (0.05)	$\text{var}(i_t, c_t)$ (0.04)
σ_c	$\text{var}(h_t, h_t)$ (0.18)	$\text{cov}(h_t, h_{t+1})$ (0.18)	$\text{var}(r_t, h_t)$ (0.08)	$\text{var}(r_t, r_t)$ (0.06)	$\text{var}(\pi_t, h_t)$ (0.06)	$\text{cov}(h_t, r_{t+1})$ (0.05)	$\text{cov}(r_t, r_{t+1})$ (0.05)
λ	$\text{var}(h_t, h_t)$ (0.19)	$\text{cov}(h_t, h_{t+1})$ (0.19)	$\text{var}(r_t, r_t)$ (0.11)	$\text{cov}(r_t, r_{t+1})$ (0.09)	$\text{var}(r_t, h_t)$ (0.06)	$\text{var}(r_t, c_t)$ (0.04)	$\text{cov}(c_t, r_{t+1})$ (0.04)
ξ_w	$\text{var}(\pi_t, \pi_t)$ (0.16)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.11)	$\text{var}(r_t, \pi_t)$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{cov}(\pi_t, h_{t+1})$ (0.05)	$\text{var}(h_t, h_t)$ (0.05)
σ_l	$\text{var}(h_t, h_t)$ (0.25)	$\text{cov}(h_t, h_{t+1})$ (0.24)	$\text{var}(r_t, h_t)$ (0.09)	$\text{cov}(h_t, r_{t+1})$ (0.05)	$\text{var}(\pi_t, \pi_t)$ (0.05)	$\text{var}(\pi_t, h_t)$ (0.05)	$\text{cov}(r_t, h_{t+1})$ (0.04)
ξ_p	$\text{var}(h_t, h_t)$ (0.28)	$\text{cov}(h_t, h_{t+1})$ (0.27)	$\text{var}(\pi_t, h_t)$ (0.11)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{cov}(\pi_t, h_{t+1})$ (0.05)	$\text{var}(w_t, w_t)$ (0.02)	$\text{var}(\pi_t, \pi_t)$ (0.02)
ι_w	$\text{var}(\pi_t, h_t)$ (0.13)	$\text{var}(h_t, h_t)$ (0.12)	$\text{cov}(h_t, h_{t+1})$ (0.11)	$\text{cov}(\pi_t, h_{t+1})$ (0.07)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{cov}(w_t, \pi_{t+1})$ (0.06)	$\text{cov}(\pi_t, r_{t+1})$ (0.05)
ι_p	$\text{var}(h_t, h_t)$ (0.23)	$\text{cov}(h_t, h_{t+1})$ (0.22)	$\text{var}(\pi_t, h_t)$ (0.12)	$\text{cov}(\pi_t, h_{t+1})$ (0.08)	$\text{cov}(\pi_t, r_{t+1})$ (0.03)	$\text{var}(r_t, \pi_t)$ (0.03)	$\text{cov}(h_t, r_{t+1})$ (0.03)
ψ	$\text{var}(h_t, h_t)$ (0.41)	$\text{cov}(h_t, h_{t+1})$ (0.40)	$\text{var}(r_t, h_t)$ (0.03)	$\text{cov}(h_t, r_{t+1})$ (0.02)	$\text{cov}(r_t, h_{t+1})$ (0.02)	$\text{var}(\pi_t, h_t)$ (0.01)	$\text{var}(\pi_t, \pi_t)$ (0.01)
Φ	$\text{var}(h_t, h_t)$ (0.20)	$\text{cov}(h_t, h_{t+1})$ (0.19)	$\text{var}(\pi_t, h_t)$ (0.11)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{cov}(\pi_t, h_{t+1})$ (0.05)	$\text{var}(y_t, y_t)$ (0.03)	$\text{var}(\pi_t, \pi_t)$ (0.03)
r_π	$\text{var}(\pi_t, \pi_t)$ (0.19)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.08)	$\text{var}(r_t, \pi_t)$ (0.08)	$\text{cov}(\pi_t, h_{t+1})$ (0.05)	$\text{var}(r_t, r_t)$ (0.05)	$\text{var}(h_t, h_t)$ (0.05)
ρ	$\text{var}(r_t, r_t)$ (0.15)	$\text{cov}(r_t, r_{t+1})$ (0.14)	$\text{var}(r_t, \pi_t)$ (0.14)	$\text{var}(\pi_t, h_t)$ (0.08)	$\text{var}(\pi_t, \pi_t)$ (0.08)	$\text{cov}(\pi_t, r_{t+1})$ (0.07)	$\text{cov}(\pi_t, h_{t+1})$ (0.05)
r_y	$\text{var}(\pi_t, h_t)$ (0.11)	$\text{var}(\pi_t, \pi_t)$ (0.11)	$\text{cov}(h_t, h_{t+1})$ (0.09)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.09)	$\text{var}(h_t, h_t)$ (0.08)	$\text{var}(r_t, \pi_t)$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)
$r_{\triangle y}$	$\text{var}(h_t, h_t)$ (0.28)	$\text{cov}(h_t, h_{t+1})$ (0.27)	$\text{var}(r_t, h_t)$ (0.11)	$\text{cov}(h_t, r_{t+1})$ (0.06)	$\text{cov}(r_t, h_{t+1})$ (0.05)	$\text{var}(r_t, r_t)$ (0.02)	$\text{cov}(r_t, r_{t+1})$ (0.02)
$\bar{\pi}$	$E(\pi_t)$ (0.48)	$E(r_t)$ (0.24)	$E(c_t)$ (0.10)	$E(y_t)$ (0.07)	$E(w_t)$ (0.05)	$E(h_t)$ (0.03)	$E(i_t)$ (0.03)
β	$E(r_t)$ (0.20)	$E(\pi_t)$ (0.12)	$\text{var}(i_t, i_t)$ (0.07)	$\text{var}(i_t, y_t)$ (0.06)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(h_t, h_{t+1})$ (0.06)	$\text{var}(\pi_t, \pi_t)$ (0.06)
$\bar{l}$	$E(\pi_t)$ (0.43)	$E(r_t)$ (0.26)	$E(w_t)$ (0.14)	$E(y_t)$ (0.09)	$E(h_t)$ (0.04)	$E(c_t)$ (0.02)	$E(i_t)$ (0.01)
γ	$E(w_t)$ (0.22)	$E(r_t)$ (0.18)	$E(\pi_t)$ (0.09)	$\text{var}(h_t, h_t)$ (0.07)	$\text{cov}(h_t, h_{t+1})$ (0.07)	$E(c_t)$ (0.06)	$E(i_t)$ (0.04)
α	$\text{var}(h_t, h_t)$ (0.39)	$\text{cov}(h_t, h_{t+1})$ (0.38)	$\text{var}(i_t, i_t)$ (0.03)	$\text{var}(r_t, h_t)$ (0.03)	$\text{var}(i_t, y_t)$ (0.02)	$\text{cov}(h_t, r_{t+1})$ (0.02)	$\text{var}(r_t, r_t)$ (0.01)
ρ_a	$\text{var}(h_t, h_t)$ (0.26)	$\text{cov}(h_t, h_{t+1})$ (0.25)	$\text{var}(r_t, h_t)$ (0.09)	$\text{cov}(h_t, r_{t+1})$ (0.07)	$\text{var}(\pi_t, h_t)$ (0.04)	$\text{cov}(r_t, h_{t+1})$ (0.03)	$\text{cov}(\pi_t, h_{t+1})$ (0.02)
ρ_b	$\text{var}(r_t, r_t)$ (0.26)	$\text{cov}(r_t, r_{t+1})$ (0.22)	$\text{cov}(c_t, r_{t+1})$ (0.07)	$\text{var}(r_t, c_t)$ (0.06)	$\text{var}(c_t, c_t)$ (0.05)	$\text{cov}(h_t, r_{t+1})$ (0.04)	$\text{var}(r_t, \pi_t)$ (0.03)
ρ_g	$\text{var}(i_t, y_t)$ (0.10)	$\text{var}(h_t, h_t)$ (0.08)	$\text{cov}(h_t, h_{t+1})$ (0.07)	$\text{var}(i_t, i_t)$ (0.06)	$\text{var}(c_t, y_t)$ (0.06)	$\text{var}(c_t, c_t)$ (0.05)	$\text{var}(r_t, r_t)$ (0.04)
ρ_I	$\text{var}(i_t, i_t)$ (0.23)	$\text{cov}(h_t, h_{t+1})$ (0.09)	$\text{var}(h_t, h_t)$ (0.09)	$\text{cov}(i_t, i_{t+1})$ (0.09)	$\text{var}(i_t, c_t)$ (0.05)	$\text{var}(c_t, c_t)$ (0.03)	$\text{var}(r_t, \pi_t)$ (0.03)
ρ_r	$\text{var}(r_t, r_t)$ (0.27)	$\text{cov}(r_t, r_{t+1})$ (0.27)	$\text{cov}(r_t, \pi_{t+1})$ (0.06)	$\text{cov}(h_t, r_{t+1})$ (0.05)	$\text{var}(\pi_t, \pi_t)$ (0.04)	$\text{cov}(\pi_t, h_{t+1})$ (0.03)	$\text{var}(r_t, h_t)$ (0.03)
ρ_p	$\text{var}(h_t, h_t)$ (0.15)	$\text{cov}(h_t, h_{t+1})$ (0.15)	$\text{var}(\pi_t, \pi_t)$ (0.14)	$\text{var}(\pi_t, h_t)$ (0.13)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.11)	$\text{cov}(\pi_t, h_{t+1})$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.05)
ρ_w	$\text{var}(\pi_t, \pi_t)$ (0.18)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.15)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(h_t, h_{t+1})$ (0.05)	$\text{cov}(h_t, \pi_{t+1})$ (0.05)	$\text{var}(r_t, \pi_t)$ (0.05)
ρ_{ga}	$\text{var}(h_t, h_t)$ (0.46)	$\text{cov}(h_t, h_{t+1})$ (0.45)	$\text{var}(y_t, y_t)$ (0.03)	$\text{var}(c_t, y_t)$ (0.01)	$\text{var}(i_t, y_t)$ (0.01)	$\text{var}(h_t, c_t)$ (0.01)	$\text{var}(h_t, y_t)$ (0.01)
μ_p	$\text{var}(h_t, h_t)$ (0.15)	$\text{cov}(h_t, h_{t+1})$ (0.14)	$\text{var}(\pi_t, \pi_t)$ (0.14)	$\text{var}(\pi_t, h_t)$ (0.13)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.11)	$\text{cov}(\pi_t, h_{t+1})$ (0.08)	$\text{cov}(h_t, \pi_{t+1})$ (0.05)
μ_w	$\text{var}(\pi_t, \pi_t)$ (0.20)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.17)	$\text{var}(\pi_t, h_t)$ (0.10)	$\text{cov}(h_t, \pi_{t+1})$ (0.06)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(h_t, h_{t+1})$ (0.06)	$\text{var}(r_t, \pi_t)$ (0.05)
σ_a	$\text{var}(h_t, h_t)$ (0.49)	$\text{cov}(h_t, h_{t+1})$ (0.48)	$\text{var}(y_t, y_t)$ (0.01)	$\text{var}(h_t, y_t)$ (0.01)	$\text{cov}(y_t, h_{t+1})$ (0.01)	$\text{var}(h_t, w_t)$ (0.00)	$\text{cov}(w_t, h_{t+1})$ (0.00)
σ_b	$\text{var}(r_t, r_t)$ (0.24)	$\text{cov}(r_t, r_{t+1})$ (0.21)	$\text{var}(c_t, c_t)$ (0.15)	$\text{var}(r_t, c_t)$ (0.09)	$\text{cov}(c_t, r_{t+1})$ (0.07)	$\text{cov}(c_t, c_{t+1})$ (0.03)	$\text{cov}(r_t, h_{t+1})$ (0.02)
σ_g	$\text{var}(h_t, h_t)$ (0.36)	$\text{cov}(h_t, h_{t+1})$ (0.35)	$\text{var}(y_t, y_t)$ (0.06)	$\text{var}(c_t, y_t)$ (0.05)	$\text{var}(i_t, y_t)$ (0.04)	$\text{var}(c_t, c_t)$ (0.03)	$\text{var}(i_t, i_t)$ (0.02)
σ_I	$\text{var}(i_t, i_t)$ (0.30)	$\text{cov}(i_t, i_{t+1})$ (0.16)	$\text{var}(h_t, h_t)$ (0.06)	$\text{cov}(h_t, h_{t+1})$ (0.06)	$\text{var}(i_t, c_t)$ (0.04)	$\text{var}(r_t, h_t)$ (0.03)	$\text{var}(r_t, i_t)$ (0.02)
σ_r	$\text{var}(r_t, r_t)$ (0.19)	$\text{cov}(r_t, r_{t+1})$ (0.17)	$\text{var}(h_t, h_t)$ (0.11)	$\text{cov}(h_t, h_{t+1})$ (0.10)	$\text{var}(r_t, h_t)$ (0.09)	$\text{var}(r_t, \pi_t)$ (0.06)	$\text{cov}(h_t, r_{t+1})$ (0.06)
σ_p	$\text{var}(\pi_t, \pi_t)$ (0.16)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.14)	$\text{var}(h_t, h_t)$ (0.14)	$\text{cov}(h_t, h_{t+1})$ (0.13)	$\text{var}(\pi_t, h_t)$ (0.13)	$\text{cov}(h_t, \pi_{t+1})$ (0.07)	$\text{cov}(\pi_t, h_{t+1})$ (0.06)
σ_w	$\text{var}(\pi_t, \pi_t)$ (0.21)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.18)	$\text{var}(w_t, w_t)$ (0.10)	$\text{var}(\pi_t, h_t)$ (0.06)	$\text{var}(r_t, \pi_t)$ (0.05)	$\text{var}(\pi_t, w_t)$ (0.04)	$\text{cov}(h_t, \pi_{t+1})$ (0.04)

Note: The GMM objective function uses moments up to first order (cross) autocovariances and optimal weighting matrix.

Table A.5: Moment with largest weights assuming no correlation

param.							
φ	$\text{cov}(r_t, r_{t+1})$ (0.04)	$\text{cov}(r_t, r_{t+2})$ (0.04)	$\text{cov}(r_t, r_{t+3})$ (0.03)	$\text{var}(i_t, i_t)$ (0.03)	$\text{cov}(i_t, i_{t+1})$ (0.02)	$\text{cov}(r_t, r_{t+4})$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)
σ_c	$E(r_t)$ (0.10)	$\text{cov}(r_t, r_{t+1})$ (0.02)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)	$\text{cov}(r_t, r_{t+3})$ (0.01)	$\text{var}(i_t, i_t)$ (0.01)	$\text{cov}(i_t, i_{t+1})$ (0.01)
λ	$\text{cov}(r_t, r_{t+1})$ (0.04)	$\text{cov}(r_t, r_{t+2})$ (0.03)	$\text{var}(r_t, r_t)$ (0.02)	$\text{cov}(r_t, r_{t+3})$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.01)	$\text{var}(c_t, c_t)$ (0.01)	$\text{var}(r_t, \pi_t)$ (0.01)
ξ_w	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+3})$ (0.02)	$\text{cov}(h_t, h_{t+4})$ (0.02)	$\text{cov}(h_t, h_{t+5})$ (0.02)	$\text{cov}(h_t, h_{t+6})$ (0.01)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.01)
σ_l	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+3})$ (0.02)	$\text{cov}(h_t, h_{t+4})$ (0.02)	$\text{cov}(h_t, h_{t+5})$ (0.02)	$\text{cov}(h_t, h_{t+6})$ (0.02)	$\text{cov}(h_t, h_{t+7})$ (0.01)
ξ_p	$\text{var}(w_t, w_t)$ (0.06)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.02)	$\text{cov}(r_t, r_{t+1})$ (0.01)	$\text{cov}(w_t, w_{t+1})$ (0.01)	$\text{cov}(r_t, \pi_{t+2})$ (0.01)	$\text{cov}(r_t, \pi_{t+1})$ (0.01)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.01)
ι_w	$\text{cov}(\pi_t, \pi_{t+1})$ (0.02)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.01)	$\text{cov}(\pi_t, \pi_t)$ (0.01)	$\text{cov}(r_t, \pi_{t+1})$ (0.01)	$\text{cov}(r_t, r_{t+1})$ (0.01)
ι_p	$\text{cov}(\pi_t, \pi_{t+1})$ (0.04)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{var}(\pi_t, \pi_t)$ (0.02)	$\text{var}(w_t, w_t)$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{cov}(r_t, \pi_{t+2})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)
ψ	$\text{var}(y_t, y_t)$ (0.02)	$\text{var}(i_t, i_t)$ (0.02)	$\text{cov}(i_t, i_{t+1})$ (0.01)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.01)	$\text{var}(i_t, y_t)$ (0.01)	$\text{var}(\pi_t, \pi_t)$ (0.01)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.01)
Φ	$\text{cov}(h_t, h_{t+1})$ (0.03)	$\text{cov}(h_t, h_{t+2})$ (0.03)	$\text{cov}(h_t, h_{t+3})$ (0.03)	$\text{cov}(h_t, h_{t+4})$ (0.03)	$\text{cov}(h_t, h_{t+5})$ (0.02)	$\text{cov}(h_t, h_{t+6})$ (0.02)	$\text{var}(h_t, h_t)$ (0.02)
r_π	$\text{cov}(\pi_t, \pi_{t+1})$ (0.03)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.03)	$\text{var}(\pi_t, \pi_t)$ (0.02)	$\text{cov}(\pi_t, \pi_{t+3})$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{cov}(r_t, \pi_{t+2})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)
ρ	$\text{cov}(r_t, r_{t+1})$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)	$\text{cov}(r_t, r_{t+3})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{cov}(r_t, r_{t+4})$ (0.01)
r_y	$\text{cov}(\pi_t, \pi_{t+1})$ (0.02)	$\text{cov}(r_t, r_{t+1})$ (0.02)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)	$\text{cov}(r_t, \pi_{t+2})$ (0.02)	$\text{cov}(r_t, r_{t+2})$ (0.02)
$r_{\triangle y}$	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.02)	$\text{var}(c_t, y_t)$ (0.01)	$\text{var}(c_t, c_t)$ (0.01)	$\text{cov}(h_t, h_{t+2})$ (0.01)	$\text{cov}(h_t, h_{t+3})$ (0.01)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.01)
$\bar{\pi}$	$E(r_t)$ (0.63)	$E(\pi_t)$ (0.37)	$E(y_t)$ (0.00)	$E(c_t)$ (0.00)	$E(i_t)$ (0.00)	$E(w_t)$ (0.00)	$E(h_t)$ (0.00)
β	$E(r_t)$ (0.50)	$\text{cov}(h_t, h_{t+1})$ (0.01)	$\text{cov}(h_t, h_{t+2})$ (0.01)	$\text{cov}(h_t, h_{t+3})$ (0.01)	$\text{cov}(r_t, r_{t+1})$ (0.01)	$\text{cov}(h_t, h_{t+4})$ (0.01)	$\text{cov}(h_t, h_{t+5})$ (0.01)
$\bar{l}$	$E(h_t)$ (1.00)	$E(y_t)$ (0.00)	$E(c_t)$ (0.00)	$E(i_t)$ (0.00)	$E(w_t)$ (0.00)	$E(\pi_t)$ (0.00)	$E(r_t)$ (0.00)
γ	$E(r_t)$ (0.49)	$E(w_t)$ (0.13)	$E(c_t)$ (0.09)	$E(y_t)$ (0.05)	$E(i_t)$ (0.01)	$\text{cov}(h_t, h_{t+1})$ (0.01)	$\text{cov}(h_t, h_{t+2})$ (0.01)
α	$\text{var}(i_t, i_t)$ (0.03)	$\text{cov}(i_t, i_{t+1})$ (0.03)	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{var}(y_t, y_t)$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.02)	$\text{cov}(r_t, r_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+3})$ (0.02)
ρ_a	$\text{var}(c_t, y_t)$ (0.03)	$\text{var}(c_t, c_t)$ (0.02)	$\text{cov}(r_t, r_{t+1})$ (0.02)	$\text{var}(y_t, y_t)$ (0.02)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{var}(w_t, w_t)$ (0.02)	$\text{var}(r_t, r_t)$ (0.01)
ρ_b	$\text{var}(c_t, c_t)$ (0.08)	$\text{var}(c_t, y_t)$ (0.07)	$\text{cov}(r_t, r_{t+1})$ (0.05)	$\text{var}(y_t, y_t)$ (0.04)	$\text{var}(r_t, r_t)$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.03)	$\text{cov}(c_t, c_{t+1})$ (0.03)
ρ_g	$\text{cov}(h_t, h_{t+1})$ (0.05)	$\text{cov}(h_t, h_{t+2})$ (0.05)	$\text{cov}(h_t, h_{t+3})$ (0.04)	$\text{cov}(h_t, h_{t+4})$ (0.04)	$\text{cov}(h_t, h_{t+5})$ (0.04)	$\text{cov}(h_t, h_{t+6})$ (0.04)	$\text{cov}(h_t, h_{t+7})$ (0.03)
ρ_I	$\text{var}(i_t, i_t)$ (0.05)	$\text{cov}(i_t, i_{t+1})$ (0.04)	$\text{cov}(r_t, r_{t+1})$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.03)	$\text{cov}(r_t, r_{t+3})$ (0.02)	$\text{var}(i_t, y_t)$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)
ρ_r	$\text{var}(c_t, c_t)$ (0.03)	$\text{var}(c_t, y_t)$ (0.03)	$\text{cov}(r_t, r_{t+1})$ (0.02)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.02)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{var}(y_t, y_t)$ (0.02)	$\text{cov}(h_t, h_{t+1})$ (0.02)
ρ_p	$\text{cov}(\pi_t, \pi_{t+1})$ (0.03)	$\text{var}(w_t, w_t)$ (0.02)	$\text{var}(\pi_t, \pi_t)$ (0.02)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.01)
ρ_w	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+3})$ (0.02)	$\text{cov}(h_t, h_{t+4})$ (0.02)	$\text{cov}(h_t, h_{t+5})$ (0.01)	$\text{cov}(h_t, h_{t+6})$ (0.01)	$\text{cov}(h_t, h_{t+7})$ (0.01)
ρ_{ga}	$\text{var}(y_t, y_t)$ (0.05)	$\text{cov}(r_t, r_{t+1})$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{cov}(r_t, r_{t+3})$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)	$\text{cov}(r_t, r_{t+4})$ (0.01)	$\text{var}(r_t, \pi_t)$ (0.01)
μ_p	$\text{cov}(\pi_t, \pi_{t+1})$ (0.03)	$\text{var}(w_t, w_t)$ (0.02)	$\text{var}(\pi_t, \pi_t)$ (0.02)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)	$\text{cov}(r_t, \pi_{t+2})$ (0.02)
μ_w	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+3})$ (0.02)	$\text{cov}(h_t, h_{t+4})$ (0.02)	$\text{cov}(h_t, h_{t+5})$ (0.01)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.01)	$\text{cov}(h_t, h_{t+6})$ (0.01)
σ_a	$\text{var}(y_t, y_t)$ (0.03)	$\text{cov}(r_t, r_{t+1})$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.03)	$\text{var}(i_t, y_t)$ (0.02)	$\text{var}(c_t, y_t)$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)	$\text{cov}(r_t, r_{t+3})$ (0.02)
σ_b	$\text{var}(c_t, c_t)$ (0.16)	$\text{var}(c_t, y_t)$ (0.12)	$\text{var}(y_t, y_t)$ (0.06)	$\text{cov}(r_t, r_{t+1})$ (0.04)	$\text{var}(r_t, r_t)$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{var}(i_t, y_t)$ (0.02)
σ_g	$\text{var}(y_t, y_t)$ (0.05)	$\text{cov}(h_t, h_{t+1})$ (0.03)	$\text{cov}(h_t, h_{t+2})$ (0.03)	$\text{cov}(h_t, h_{t+3})$ (0.03)	$\text{cov}(h_t, h_{t+4})$ (0.02)	$\text{cov}(h_t, h_{t+5})$ (0.02)	$\text{cov}(h_t, h_{t+6})$ (0.02)
σ_I	$\text{var}(i_t, i_t)$ (0.08)	$\text{cov}(i_t, i_{t+1})$ (0.05)	$\text{var}(i_t, y_t)$ (0.04)	$\text{cov}(r_t, r_{t+1})$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{cov}(r_t, r_{t+3})$ (0.02)	$\text{var}(r_t, r_t)$ (0.02)
σ_r	$\text{cov}(r_t, r_{t+1})$ (0.04)	$\text{var}(r_t, r_t)$ (0.04)	$\text{var}(c_t, c_t)$ (0.03)	$\text{var}(c_t, y_t)$ (0.03)	$\text{cov}(r_t, r_{t+2})$ (0.02)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)
σ_p	$\text{var}(w_t, w_t)$ (0.04)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.03)	$\text{var}(\pi_t, \pi_t)$ (0.03)	$\text{cov}(r_t, \pi_{t+1})$ (0.02)	$\text{var}(r_t, \pi_t)$ (0.02)	$\text{cov}(\pi_t, \pi_{t+2})$ (0.02)	$\text{cov}(r_t, \pi_{t+2})$ (0.02)
σ_w	$\text{cov}(h_t, h_{t+1})$ (0.02)	$\text{cov}(h_t, h_{t+2})$ (0.02)	$\text{cov}(h_t, h_{t+3})$ (0.02)	$\text{cov}(h_t, h_{t+4})$ (0.02)	$\text{cov}(h_t, h_{t+5})$ (0.01)	$\text{cov}(\pi_t, \pi_{t+1})$ (0.01)	$\text{cov}(h_t, h_{t+6})$ (0.01)

Note: Based on the weights assigned in the score when the weighting matrix is diagonal.

Table A.6: IM decomposition the posterior mean

	CRLB	sens.	coll.	$\boldsymbol{\varrho}_i$	$\boldsymbol{\varrho}_{i(1)}$	$\boldsymbol{\varrho}_{i(2)}$	$\boldsymbol{\varrho}_{i(3)}$	$\boldsymbol{\varrho}_{i(4)}$
φ	2.350	0.964	2.437	0.912	0.718 (ρ_I)	0.824 (λ,ρ_I)	0.833 (λ,Φ,ρ_I)	0.854 (λ,ξ_w,ρ_I,λ_w)
σ_c	0.295	0.062	4.726	0.977	0.735 (λ)	0.836 (ξ_w,λ_w)	0.886 (λ,ξ_w,λ_w)	0.931 (β,λ,ξ_w,λ_w)
λ	0.064	0.020	3.211	0.950	0.735 (σ_c)	0.801 (σ_c,$r_{\triangle y}$)	0.826 (σ_c,$r_{\triangle y}$,ρ_b)	0.857 (σ_c,σ_l,$r_{\triangle y}$,λ_w)
ξ_w	0.188	0.017	11.051	0.996	0.946 (λ_w)	0.981 (σ_c,λ_w)	0.987 (μ_w,σ_c,λ_w)	0.989 (β,μ_w,σ_c,λ_w)
σ_l	1.014	0.411	2.470	0.914	0.783 (λ_w)	0.813 (λ,λ_w)	0.849 (λ,$r_{\triangle y}$,λ_w)	0.863 (μ_w,λ,$r_{\triangle y}$,λ_w)
ξ_p	0.059	0.023	2.640	0.926	0.796 (ρ_p)	0.828 (ξ_w,ρ_p)	0.853 (ξ_w,ρ_p,σ_w)	0.872 (ξ_w,r_π,ρ_p,σ_w)
ι_w	0.206	0.144	1.435	0.717	0.440 (σ_w)	0.573 (σ_p,σ_w)	0.653 (ξ_w,σ_p,σ_w)	0.687 (ξ_w,ρ_p,σ_p,σ_w)
ι_p	0.131	0.058	2.276	0.898	0.813 (μ_p)	0.851 (μ_p,σ_p)	0.867 (μ_p,ρ_p,σ_p)	0.881 (μ_w,μ_p,ρ_p,σ_p)
ψ	0.171	0.099	1.729	0.816	0.528 (δ)	0.569 ($r_{\triangle y}$,δ)	0.608 (λ,$r_{\triangle y}$,δ)	0.643 (Φ,ξ_p,σ_a,δ)
Φ	0.136	0.064	2.133	0.883	0.461 (ξ_p)	0.559 (ξ_p,σ_a)	0.631 (α,ξ_p,σ_a)	0.671 (α,ξ_p,σ_a,δ)
r_π	0.392	0.121	3.236	0.951	0.566 (ρ)	0.866 (r_y,ρ)	0.910 (σ_c,r_y,ρ)	0.922 (σ_c,r_y,ρ,ρ_I)
ρ	0.044	0.016	2.803	0.934	0.566 (r_π)	0.825 (r_π,r_y)	0.868 (σ_c,r_π,r_y)	0.895 (σ_c,r_π,r_y,ρ_r)
r_y	0.038	0.014	2.762	0.932	0.508 (r_π)	0.807 (r_π,ρ)	0.882 (σ_c,r_π,ρ)	0.897 (σ_c,r_π,ρ,ρ_I)
$r_{\triangle y}$	0.044	0.022	2.018	0.869	0.593 (λ)	0.648 (ψ,λ)	0.691 (λ,σ_l,λ_w)	0.738 (ψ,λ,σ_l,λ_w)
$\bar{\pi}$	0.227	0.130	1.745	0.820	0.811 ($\bar{l}$)	0.817 ($\bar{l}$,β)	0.819 ($\bar{l}$,β,γ)	0.819 ($\bar{l}$,β,α,γ)
$100\left(\beta^{-1}-1\right)$	0.160	0.088	1.831	0.838	0.364 ($\bar{\pi}$)	0.436 ($\bar{\pi}$,α)	0.525 (σ_c,ξ_w,λ_w)	0.623 (σ_c,λ,ξ_w,λ_w)
$\bar{l}$	1.486	0.859	1.730	0.816	0.811 ($\bar{\pi}$)	0.815 ($\bar{\pi}$,γ)	0.816 ($\bar{\pi}$,β,γ)	0.816 ($\bar{\pi}$,β,α,γ)
γ	0.010	0.010	1.016	0.176	0.111 ($\bar{l}$)	0.145 ($\bar{l}$,$\bar{\pi}$)	0.173 ($\bar{l}$,$\bar{\pi}$,β)	0.175 ($\bar{l}$,$\bar{\pi}$,β,α)
α	0.022	0.013	1.630	0.790	0.595 (δ)	0.656 (Φ,δ)	0.685 (Φ,$r_{\triangle y}$,δ)	0.712 (Φ,$r_{\triangle y}$,σ_a,δ)
ρ_a	0.016	0.010	1.568	0.770	0.355 (ρ_g)	0.530 (ρ_g,δ)	0.643 (σ_c,ρ_g,δ)	0.664 (σ_c,λ,ρ_g,δ)
ρ_b	0.089	0.042	2.143	0.885	0.827 (σ_b)	0.869 (λ,σ_b)	0.871 (σ_c,λ,σ_b)	0.872 (λ,ξ_w,σ_b,λ_w)
ρ_g	0.011	0.008	1.363	0.680	0.355 (ρ_a)	0.446 (ρ_a,δ)	0.548 (σ_c,ρ_a,δ)	0.570 (σ_c,λ,ρ_a,δ)
ρ_I	0.066	0.028	2.350	0.905	0.840 (σ_I)	0.877 (φ,σ_I)	0.884 (φ,$r_{\triangle y}$,σ_I)	0.888 (φ,$r_{\triangle y}$,σ_I,δ)
ρ_r	0.093	0.072	1.295	0.636	0.460 (ρ)	0.557 ($r_{\triangle y}$,ρ)	0.574 ($r_{\triangle y}$,ρ,σ_r)	0.584 ($r_{\triangle y}$,ρ,ρ_b,σ_r)
ρ_p	0.058	0.013	4.441	0.974	0.963 (μ_p)	0.968 (μ_p,ξ_p)	0.969 (μ_p,ι_p,ξ_p)	0.973 (μ_p,ι_p,ξ_p,σ_p)
ρ_w	0.016	0.008	2.036	0.871	0.761 (ξ_w)	0.812 (ξ_w,δ)	0.824 (ξ_w,r_y,δ)	0.830 (μ_w,ξ_w,r_y,δ)
ρ_{ga}	0.103	0.092	1.114	0.441	0.233 (Φ)	0.263 (Φ,ξ_p)	0.290 (Φ,ξ_p,g_y)	0.313 (Φ,ξ_p,r_y,g_y)
μ_p	0.148	0.023	6.441	0.988	0.963 (ρ_p)	0.976 (ρ_p,σ_p)	0.986 (ι_p,ρ_p,σ_p)	0.987 (μ_w,ι_p,ρ_p,σ_p)
μ_w	0.059	0.016	3.712	0.963	0.906 (ξ_w)	0.935 (ξ_w,σ_w)	0.942 (ξ_w,r_π,σ_w)	0.947 (ξ_w,σ_l,r_π,σ_w)
σ_a	0.035	0.026	1.345	0.669	0.316 (Φ)	0.377 (ψ,Φ)	0.433 (ψ,Φ,ξ_p)	0.486 (α,ψ,Φ,δ)
σ_b	0.027	0.014	1.945	0.858	0.827 (ρ_b)	0.836 (λ,ρ_b)	0.842 (σ_c,λ,ρ_b)	0.844 (σ_c,λ,ρ_b,g_y)
σ_g	0.035	0.030	1.168	0.517	0.215 (Φ)	0.262 (ψ,Φ)	0.329 (ψ,Φ,g_y)	0.372 (ψ,Φ,ξ_p,g_y)
σ_I	0.051	0.026	1.973	0.862	0.840 (ρ_I)	0.842 (ρ_I,δ)	0.843 (α,ρ_I,δ)	0.846 (σ_l,ρ_I,ρ_w,δ)
σ_r	0.016	0.014	1.137	0.476	0.294 ($r_{\triangle y}$)	0.344 (r_π,$r_{\triangle y}$)	0.371 (λ,r_π,$r_{\triangle y}$)	0.403 (r_π,$r_{\triangle y}$,r_y,ρ_r)
σ_p	0.022	0.008	2.810	0.935	0.880 (μ_p)	0.904 (μ_p,ι_p)	0.918 (μ_p,ι_w,ι_p)	0.925 (μ_p,ι_w,ι_p,ρ_p)
σ_w	0.028	0.014	1.994	0.865	0.725 (μ_w)	0.803 (μ_w,ι_w)	0.819 (μ_w,ι_w,ξ_p)	0.827 (μ_w,ι_w,ξ_p,ρ_w)
δ	0.010	0.005	2.236	0.894	0.595 (α)	0.710 (α,g_y)	0.757 (α,ρ_a,g_y)	0.785 (α,ψ,ρ_a,g_y)
λ_w	0.820	0.073	11.220	0.996	0.946 (ξ_w)	0.984 (σ_c,ξ_w)	0.988 (β,σ_c,ξ_w)	0.991 (β,φ,σ_c,ξ_w)
g_y	0.071	0.046	1.554	0.766	0.458 (δ)	0.543 ($r_{\triangle y}$,δ)	0.588 ($r_{\triangle y}$,r_y,δ)	0.604 (Φ,$r_{\triangle y}$,r_y,δ)

Note: see note to Table 2. The difference is that here δ, λ_w and g_y are free parameters.

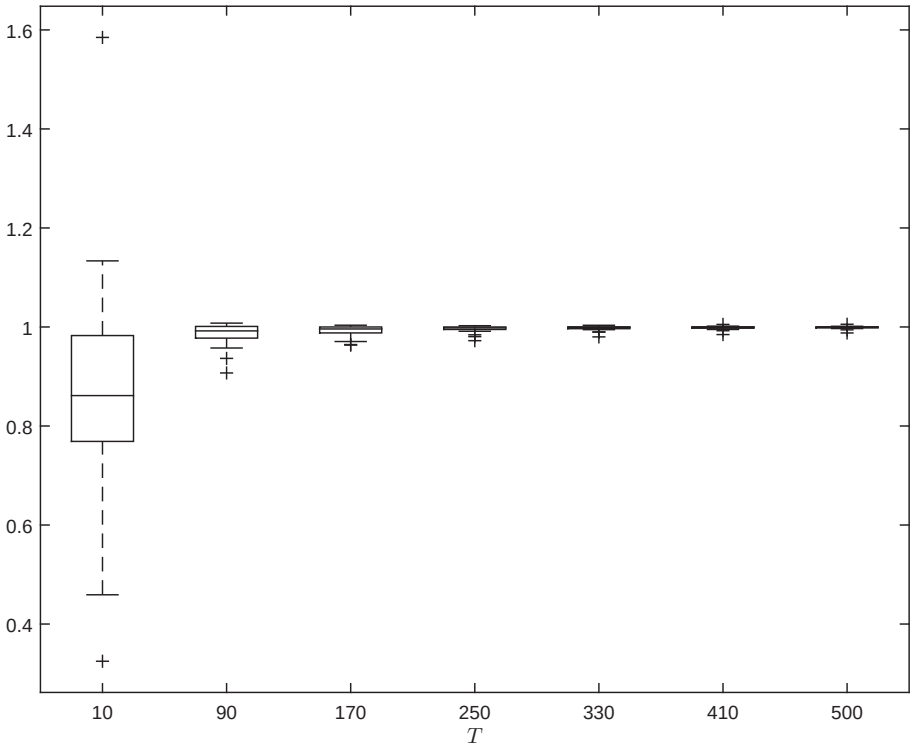

Figure A.1: Ratios of frequency domain to exact Cramér-Rao lower bounds

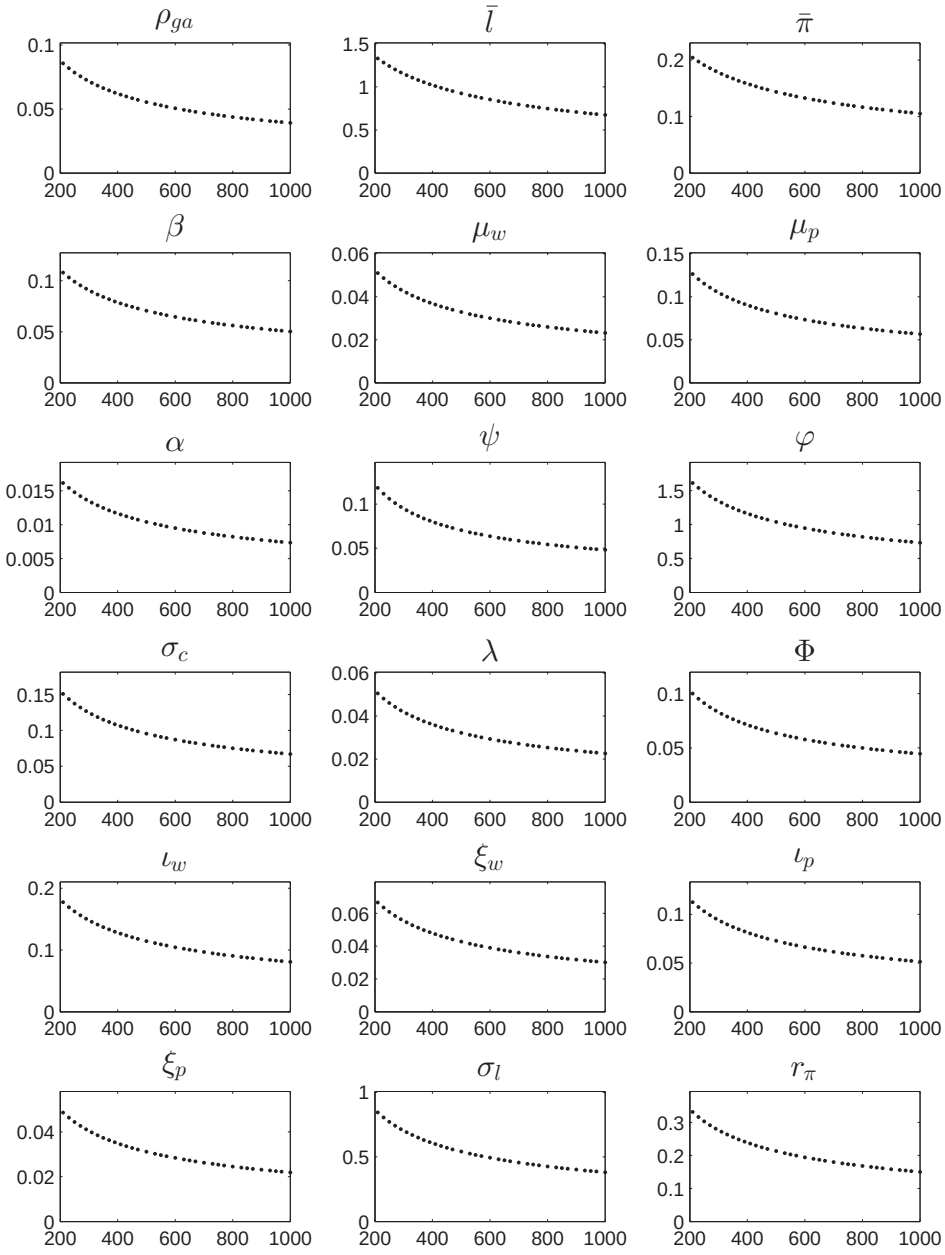

Figure A.2: Cramér-Rao lower bounds for sample sizes between 200 and 1000.

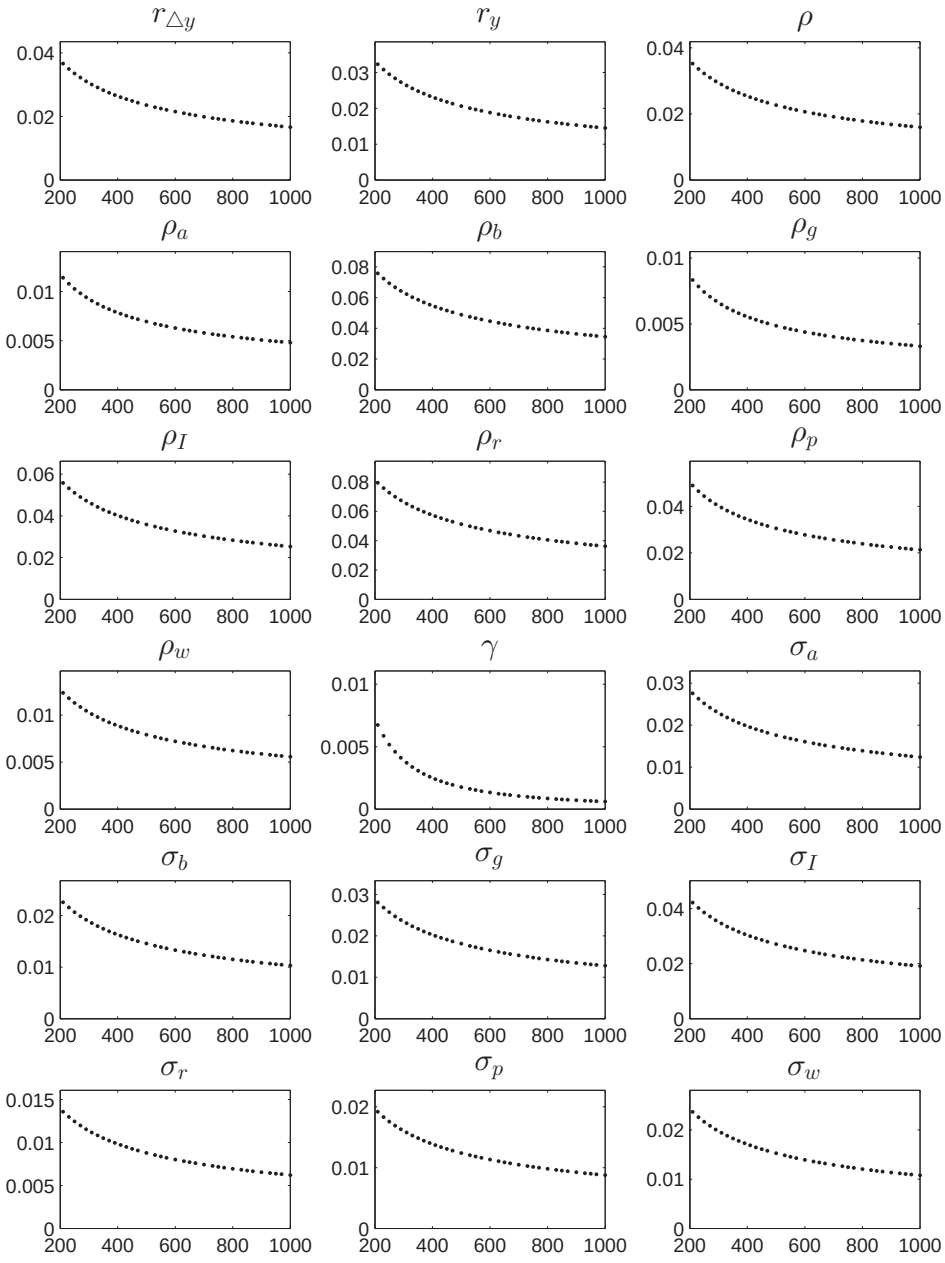

Figure A.3: Cramér-Rao lower bounds for sample sizes between 200 and 1000.

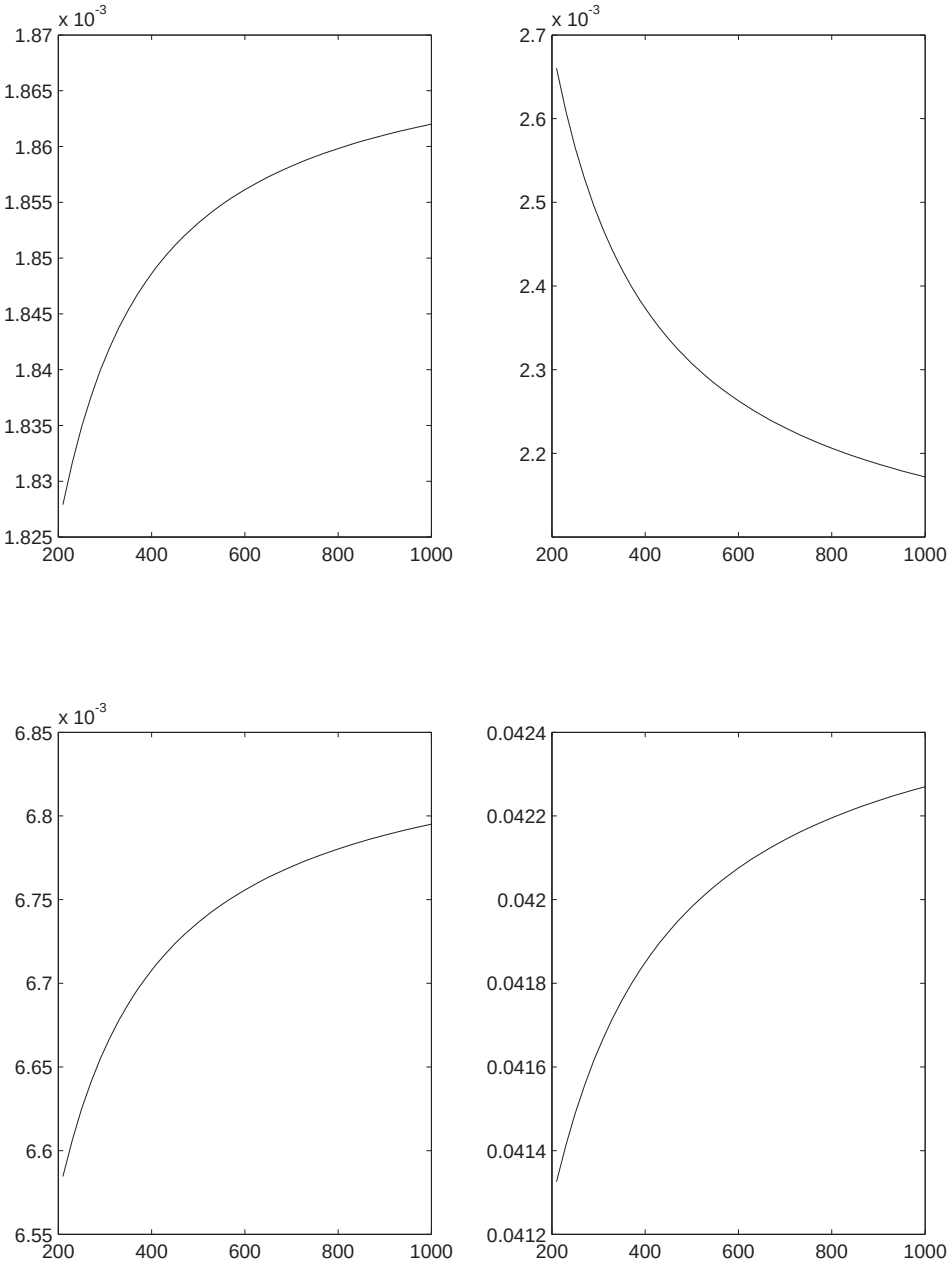

Figure A.4: Smallest eigenvalues of $\frac{1}{T}\boldsymbol{\mathcal{I}}_T$ for sample sizes between 200 and 1000.

References

Altug, S. (1989): "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30, 889–920.

Andrews, D. W. and J. H. Stock (2005): "Inference with Weak Instruments," Cowles Foundation Discussion Papers 1530, Cowles Foundation, Yale University.

Andrews, I. and A. Mikusheva (2014): "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Quantitative Economics*.

Bekker, P. A. and D. S. G. Pollock (1986): "Identification of linear stochastic models with covariance restrictions," *Journal of Econometrics*, 31, 179–208, available at http://ideas.repec.org/a/eee/econom/v31y1986i2p179-208.html.

Beyer, A. and R. E. A. Farmer (2004): "On the indeterminacy of New-Keynesian economics," Working Paper Series 323, European Central Bank, available at http://ideas.repec.org/p/ecb/ecbwps/20040323.html.

Bowden, R. J. (1973): "The Theory of Parametric Identification," *Econometrica*, 41, 1069–74.

Canova, F. and L. Sala (2009): "Back to square one: identification issues in DSGE models," *Journal of Monetary Economics, forthcoming*.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2009): "New Keynesian Models: Not Yet Useful for Policy Analysis," *American Economic Journal: Macroeconomics*, 1, 242–66.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45, available at http://ideas.repec.org/a/ucp/jpolec/v113y2005i1p1-45.html.

Christiano, L. J. and R. J. Vigfusson (2003): "Maximum likelihood in the frequency domain: the importance of time-to-plan," *Journal of Monetary Economics*, 50, 789–815.

Cochrane, J. H. (2007): "Identification with Taylor Rules: A Critical Review," NBER Working Papers 13410, National Bureau of Economic Research, Inc.

Davies, R. B. (1983): *Optimal inference in the frequency domain*, 73–92.

Diebold, F. X., L. E. Ohanian, and J. Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies*, 65, 433–51.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist inference in weakly identified dynamic stochastic general equilibrium models," *Quantitative Economics*, 4, 197–229.

Guerron-Quintana, P. A. (2010): "What you match does matter: the effects of data on DSGE estimation," *Journal of Applied Econometrics*, 25, 774–804.

Hansen, L. P. and T. J. Sargent (2014): *Recursive Models of Dynamic Linear Economies*, Princeton University Press.

Inoue, A. and B. Rossi (2011): "Testing for weak identification in possibly nonlinear models," *Journal of Econometrics*, 161, 246–261.

Iskrev, N. (2008): "Evaluating the information matrix in linearized DSGE models," *Economics Letters*, 99, 607–610.

——— (2010): "Local identification in DSGE models," *Journal of Monetary Economics*, 57, 189–202.

——— (2014): "On the distibution of information in the moment structure of DSGE models," Unpublished manuscript.

Kim, J. (2003): "Functional equivalence between intertemporal and multisectoral investment adjustment costs," *Journal of Economic Dynamics and Control*, 27, 533–549.

Klein, A. and H. Neudecker (2000): "A direct derivation of the exact Fisher information matrix of Gaussian vector state space models," *Linear Algebra and its Applications*, 321, 233–238.

Komunjer, I. and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79, 1995–2032.

Koop, G., M. H. Pesaran, and R. P. Smith (2013): "On Identification of Bayesian DSGE Models," *Journal of Business & Economic Statistics*, 31, 300–314.

Koopmans, T. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17, 125–144.

Magnus, J. R. and K. M. Abadir (2005): *Matrix Algebra*, Cambridge University Press.

Manski, C. F. (1995): *Identification Problems in Social Sciences*, Harvard University Press.

Pesaran, M. H. (1989): *The Limits to Rational Expectations*, Basil Blackwell, Oxford.

Qu, Z. (2014): "Inference in dynamic stochastic general equilibrium models with possible weak identification," *Quantitative Economics*, 5, 457–494.

Qu, Z. and D. Tkachenko (2012a): *Frequency Domain Analysis of Medium Scale DSGE Models with Application to Smets and Wouters (2007)*, chap. 13, 319–385.

——— (2012b): "Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models," *Quantitative Economics*, 3, 95–132.

Rao, C. R. (1973): *Linear Statistical Inference and its Applications*, Wiley, New York.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–91, available at http://ideas.repec.org/a/ecm/emetrp/v39y1971i3p577-91.html.

Sala, L. (2014): "DSGE MODELS IN THE FREQUENCY DOMAIN," *Journal of Applied Econometrics*.

Sargent, T. J. (1976): "The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics," *Journal of Political Economy*, 84, 631–40, available at http://ideas.repec.org/a/ucp/jpolec/v84y1976i3p631-40.html.

Segal, M. and E. Weinstein (1989): "A new method for evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems." *IEEE Transactions on Information Theory*, 35, 682–.

Shapiro, A. and M. Browne (1983): "On the investigation of local identifiability: A counterexample," *Psychometrika*, 48, 303–304.

Smets, F. and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175, available at http://ideas.repec.org/a/tpr/jeurec/v1y2003i5p1123-1175.html.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 97, 586–606.

Stock, J. H. and J. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096, available at http://ideas.repec.org/a/ecm/emetrp/v68y2000i5p1055-1096.html.

Stock, J. H. and M. Yogo (2005): *Testing for weak instruments in linear IV regression*, Cambridge University Press, chap. 5.

Tucker, L., L. Cooper, and W. Meredith (1972): "Obtaining squared multiple correlations from a correlation matrix which may be singular," *Psychometrika*, 37, 143–148.

Wansbeek, T. and E. Meijer (2000): *Measurement Error and Latent Variables in Econometrics*, Advanced Textbooks in Economics Series, Elsevier Science Limited.

Whittle, P. (1953): "The Analysis of Multiple Stationary Time Series," *Journal of the Royal Statistical Society. Series B (Methodological)*, 15, pp. 125–139.

Zadrozny, P. and S. Mittnik (1994): "Kalman-filtering methods for computing information matrices for time-invariant, periodic, and generally time-varying VARMA models and samples," *Computers and Mathematics with Applications*, 28, 107 – 119.

Zadrozny, P. A. (1989): "Analytic derivatives for estimation of linear dynamic models," *Computers and Mathematics with Applications*, 18, 539–553.

Zeira, A. and A. Nehorai (1990): "Frequency domain Cramer-Rao bound for Gaussian processes," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38, 1063–1066.